Adpositional grammars: a multilingual grammar formalism for NLP

Federico Gobbo

Adpositional Grammars

A multilingual grammar formalism for NLP

PHD DISSERTATION

Supervisor: Gaetano A. Lanzarone

Internal Advisor: Marco Benini

External Advisor: Fabrizio A. Pennacchietti

Reviewers:

Cem Bozsahin, Wim Jansen,
Flavio De Paoli, Mario Ornaghi

Dipartimento di Informatica e Comunicazione
Università degli Studi dell'Insubria
2009, January

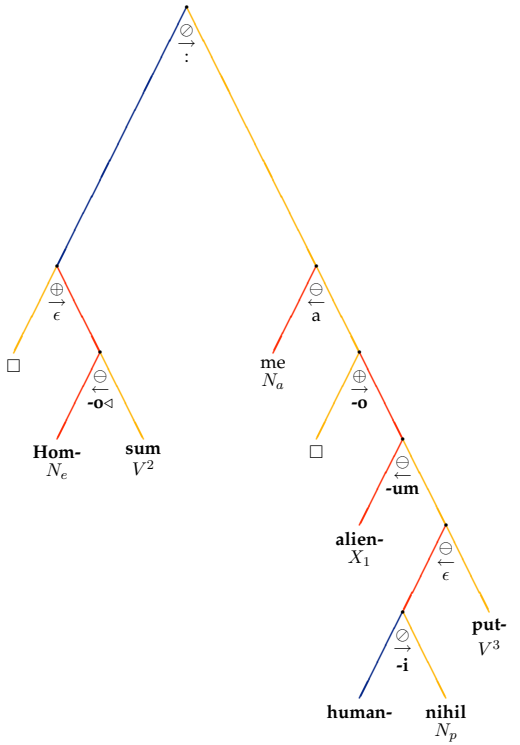


Figure 1: Publius Terentius Afer, *Heauton Timorumenos*, 77.

Typographical conventions

This dissertation is written in English and adopts the British spelling (however, Italians and Esperantophones can take advantage of the Appendix). The font is Palatino. Linguistic forms are printed in italics: *chess*. Unacceptable forms are marked by a star: **two chess player*. Uncertain forms are marked by a question mark: ?*two chess actors*. Meanings of linguistic forms, and glosses of non-English language forms, are given between single quotes: *scacchi* in Italian means 'chess', and it is a *very* important game in A.I. history. I consider English as a vacuum cleaner of lexicon, so 'foreign words' will never be in italics.

Inline citations are marked by double quotes. Example: Kasparov declared: "chess is mental torture".[1] Longer citations are put in paragraphs with smaller characters:

> I have to play an opponent, a very powerful opponent, that studied all my games, that has a unique ability – the best on the planet – to collect all this information and analyze that, and I know nothing about him... I said 'him.' I meant 'it.'[2]

Names of semantic categories and moduli are printed in small capitals: CHESS. Semantic features are printed in small capitals enclosed in square brackets: CHESS is [-ANIMATE] and [±OBJECT], according to context – i.e., the board is [+OBJECT], the game is [-OBJECT]. The implementations of moduli, i.e., the tags, are printed in sans serif: WordSeparator. Every other generic formal object is printed in typewriting: `MuteSeparator`.

Graphemes are enclosed in brace brackets {ch}. There is no use of phonetic transcription, as it is out of the scope of this dissertation. Sentences in natural languages that do not use Latin alphabet are transliterated according to the most-used convention. For instance, Chinese Mandarin excerpts are romanised in Hanyu Pinyin, while Japanese ones in Hepburn.

Bold-face is used for the first appearance of technical terms, where they are explained. Analogously, the first appearance is referred in the Index in bold-face. Further typographical conventions, strictly linked with the content expressed through the dissertation, are explained within their appropriate context.

Notes

[1] This citation is reported in Wikiquote: http://en.wikiquote.org/wiki/Chess. Retrieved 21 May 2008.

[2] Source: http://www.cnn.com/WORLD/9705/03/chess.rematch/index.html. Retrieved 21 May 2008.

Contents

List of Figures

List of Tables

Part I

Theory

Chapter 1

Introduction

How are the laws of human thought made? How can linguistic knowledge be formalised? These two questions were posed as two sides of the same coin first by Gottfried Wilhelm Leibniz in the 17^{th} century.

His answer to the first question was the *characteristica universalis*. After him, improvements were achieved by many others, among them George Boole with his calculus of logic, Gottlob Frege with his *Begriffsschrift*, Giuseppe Peano with his axiomatisation of arithmetics, not to mention Kurt Gödel's theorem and the Church-Turing thesis, which is eventually the foundation of Computer Science.[1]

In order to answer the second question, Leibniz proposed the *lingua generalis*.[2] After him, this research path was also followed by Ludwik Lejzer Zamenhof Esperanto and again by Giuseppe Peano with his *Latino sine Flexione*. Zamenhof succeeded where Peano failed, i.e., to build a stable speech community for the newborn language. As a scientific object of study, Esperanto indirectly influenced linguistics since its very beginning. In fact, the mathematician René de Saussure was also a pioneer of Esperantology, – i.e., the linguistic analysis of Esperanto – deeply influencing his brother Ferdinand, exactly during the years of his conception of linguistics and structuralism as sciences.[3] After Ferdinand de Saussure, improvements were achieved by Noam Chomsky's generative grammars and Lucien Tesnière's *syntaxe structurale*.[4]

Peano's life and work, as well as Chomsky-Schützenberger's hierarchy of formal grammars, both demonstrate that the intuition of Leibniz was correct – i.e., that these two questions are intertwined one with the other. Standing on the shoulders of giants, the aim of this dissertation is to take a step forward in this direction, which will be hopefully made by using the specific tool of adpositional grammars.

What are adpositional grammars?

Adpositional grammars (**adgrams**) are a novel grammar formalism that aims to give a general, cross-linguistical description of how human beings organise their linguistic mental spaces through the election of one or another particular morphological and syntactic construction. Hence, adgrams deal with morphology, syntax and semantics as well (in this dissertation, I will not deal with phonology, as I will take as linguistic data only written texts). Adgrams are a highly lexicalised approach to natural language (NL) analysis. I say 'adgrams' instead of 'adgram' as each NL system has its own autonomous adpositional grammar system. Here, NLs are considered synchronically, i.e., there is no treatment of how they develop and evolve.

The distinctive characteristic of adpositional grammars is that they are based on **adpositions**. For the moment, let us consider the term adposition merely as a hypernym of prepositions, postpositions and circumpositions, depending on the NL. For example, English or Italian have mainly prepositions while Japanese or Turkish have mainly postpositions. Thus, unlike other grammar formalisms, which are based on a single NL and then adapted to others, adgrams are a cross-linguistic model since the beginning.

The linguistic part presented in this dissertation is derived from research in the field of adposition by Fabrizio A. Pennacchietti for over 30 years.[5] His explorations dealt mainly with prepositions of western and semitic NLs, and were often published in journals for semitists, which are difficult to access for the non-specialist.[6] Moreover, the model was refined over the years, without reaching a systematic presentation which is valid in general, i.e., for every NL. The present work aims to describe such a model in a strong formal way. Particular attention has been given to the definitions of the technical linguistic terms, as they have been used in a rather peculiar way. Readers proficient in linguistics might be annoyed at such precision, even pedantic; please be patient, this dissertation should be readable by computer scientists, who are not familiar with linguistic technicalities. The same advice applies to computer scientists when I introduce the computational formalisms: they can be elementary to computer people, but they are not taken for granted by linguists.

The dissertation is structured as follows. In the first part, the general model of adgrams is presented, in the second and third parts a concrete instance of it will be given. Therefore, the first part is more oriented to the philosophical and linguistic aspects of adgrams, while afterwards the formal model is presented in a linguistic instance. While the general framework is derived from Pennacchietti's work, the formal model is entirely mine, developed while working closely with Marco Benini. I have chosen the Esperanto language as the first instance of adgrams, for many reasons which I will discuss later. The second part also deals with a machine trans-

lation scenario, as I think that a reliable machine translation is the best way to prove formality and cross-lingual validity of the underlying language models.

Pennacchietti has enriched and refined its model extracting concepts and data from very different sources: his great merit has been able to adapt tools and structures from everywhere so to build the adgram kernel, i.e., the analysis of each NL prepositional space (see below for details). I have grouped these sources in three major schools of linguistic research: (i) the Chomskyan school, (ii) the cognitive linguistic school, and (iii) various influences from structuralism and classic authors, in particular the school founded on Lucien Tesnière – the so-called dependency grammar school. This distinction does not mean that they do not have anything in common; I have grouped the various authors in these broad categories because I think it is the best way to present the literature under adpositional grammars.[7] The rest of this chapter is devoted to clarifying how these sources influence adgrams, before presenting the adpositional grammar model.

A final proviso is needed before going on: all linguistic data and formalisations are mine, and any mistake or error is mine only. Hereafter, ideas may or may not be shared by Pennacchietti, Benini, or whoever.

Adpositional grammars are formal

Each adgram encapsulates the language model in a self-contained system, i.e., a device which defines the set of well-formed sentences which constitute the NL. Most probably, grammar becomes automatised in learning through frequency of (un)successful use of linguistic patterns, so to build syntagms and paradigms through contrastive collocation (see details below about these terms); however, this dissertation does not deal with issues about NL acquisition, so NLs are fictitiously considered as already acquired. In this sense, adgrams pay credit to Chomsky's original purposes, as described in *Syntactic Structures* [Chomsky, 1957]: a language model can be formalised independently to the purposes for which speakers use NLs. Therefore, adgrams can be compared to other formal grammars, such as Combinatory Categorial Grammars (CCG), Tree-Adjoining Grammars (TAG) and Head-driven Phrase Structure Grammar (HPSG).[8] Adgrams aim to specify with maximum precision principles and rules which generate all and only the grammatical sentences of a NL. From a philosophical perspective, I agree with Chomsky about the belief that there exists an inbuilt structure in the mind which constrains linguistic variability. Adgrams aim exactly to describe this kind of constrictions. An immediate corollary of what said is that here NLs are neither approached as observed behaviours nor investigated in their sociolinguistic dimensions.[9] Nevertheless, rich linguistic data used in this dissertation will be extracted from real corpora – in particular the multilingual parallel corpus of the international

newspaper articles published in *Le Monde Diplomatique*.[10]

On the other hand, there are some assumptions in the Chomskyan tradition that have no place in this approach. First, the primitive categories of functions normally used in the chomskyan tradition, e.g., S, N, NP, PP, etc., are not considered valid in adgrams, as they have no immediate linguistic concretisation in terms of morphemes.[11]

Chomskyan linguistics has become more and more abstract over time, as the postulated entities and processes that constitute grammar trees became more and more theoretic: for instance, the dichotomy surface/deep level, the X-bar theory, etc., are complex structures very far from the morphemic reality. In contrast, adgrams aim to strictly adhere to the morphological entities that are visible in linguistic productions. I will refer to this principle as the **principle of linguistic adherence**. Adgrams will give account directly of the logic underlying the production of morphemes in a sentence and their collocation (see below for details). As an immediate corollary, entities like 'traces' and 'empty nodes' have been kept to a minimum: only well-known linguistic phenomena as anaphora resolution or *wh*-movements will be used. In NLs like English or Italian, these phenomena are marked by a different collocation pattern or a specific morpheme devoted to mark the phenomenon itself. I am deeply convinced that linguists cannot claim that general rules govern linguistic structures while idiosyncratic and anomalous patterns are relegated to the periphery of the system, as 'exceptions'. Adgrams give a highly general linguistic model which also gives very precise account of nuances and 'strange' linguistic patterns. More specifically, adpositional grammar trees (**adtrees**) give an account of linguistic phenomena often relegated to the 'periphery' of language in the approaches based on chomskyan linguistics. Therefore, as the reader will see from chapter 2 onwards, adtrees are quite unlike any other linguistic tree published until now.

Adpositional grammars are computational

Since this is a computer science dissertation, and since it treats linguistic topics, it falls under the rubric of 'computational linguistics', and its subclass 'natural language processing' (NLP). I am deeply convinced that natural language engineering is, or should be, a testbed of general linguistic theory and formal NL descriptions – in brief, **language models**. In my view, language models should have cross-linguistic validity and should always be tested with a formal model that can be run on a computer – perhaps it is the only testbed we have. It is important to note by now, that computational efficiency of the language model is not a theme in this dissertation. Cognitive linguists have a point claiming that for the most part our linguistic production is rarely compositional, i.e., the output of rule-based computation (this is called the **rule/list fallacy**).[12] We probably store

in our mind well-practiced patterns of use both at morphosyntactic and semantic levels. We learn to play an instrument or to drive in a similar way: initially we have to apply consciously rules and our performance is controlled, slow, and full of errors. With practice, the sequence of applied rules become automatised, and performance becomes rapid and far less prone to errors. Something similar happens in learning NLs: we do not always need to apply fine-grained rules, as we automatise established patterns that we store and recall, that indubitably fasten computation time, i.e., they increase efficiency. This means that the rules per se can be computationally inefficient. For the purposes of this dissertation, adgrams describe fine-grained rules as if there is no storage and recall of results. The goal here is to demonstrate that a concrete adgram, which is an instance of the language model, is formal, computable, and linguistically feasible.

Another important limitation of the current model of adgrams is quoting and name-entity recognition, which I see as two faces of the same coin. Adgrams cannot "understand" sentences like *Paris is a five letter word*, nor correctly translate into Italian as *Parigi è un parola di sei lettere* (in Italian 'sei' is 'six'). Analogously, the English sentence *Green Day don't like Bush* is completely out of scope, unless the name-entities 'Green Day' and 'Bush' are correctly tagged a priori – e.g., about which US president are we referring to? This kind of sentences need a lot of encyclopædic knowledge to be inserted into the machine mostly a priori, and adgrams do not compute common sense knowledge, but only linguistic-centred one (see later for details).

The formal model is modelled on a Von Neumann's machine with an intrinsical non-deterministic parser. Non-determinism is simulated by the backtracking primitive well-known in Prolog, although the formalism is closer to the so-called logical frameworks, e.g., Isabelle [Paulson, 1990]. A fully-developed abstract machine is built on the primitive abstract machine. The lexical analyser parse the text in input and gives the appropriate adpositional tree(s) in output. See chapter 7 for details.

Last but not least, adgrams are in debt with research made in the 1960s in the *Centro di Cibernetica e di Attività Linguistiche* (Centre of Cybernetics and Linguistic Activities) lead by Silvio Ceccato, in particular the formalism called Correlational Grammar (see details in chapter 2).

Adpositional grammars are cognitive

The term 'cognitive' is not free from controversy in linguistics. If we take the philosophical position of **nominalism** in its extreme form, linguistic entities are merely a matter of linguistic convention: the set of real-world entities which may be called 'dogs', or the colour values that are described as 'red' in English, have nothing in common with their name. According to nominalism, there is no intersection between linguistics and cognitive science. Note that the original position of Ferdinand de Saussure, the father of modern linguistics and structuralism, is essentially nominalist: for him, the language system (*langue*) is a system of signs where their only meaning (*sens*) is their sound patterns (*image acoustique*).[13] If the position 'using a word appropriately is simply an internal linguistic fact' were true, Eliza, Weizenbaum's famous computer program that mimics a natural-language dialogue between a Rogersian psychologist and *its* patient, would be intelligent as a human being.[14]

At the other extreme, linguistic entities are merely instances of preexisting categories like DOG and RED. In other words, categories exist independently of NLs and their users. This position is called **realism**. The fallacy of realism is revealed by the argument of the **multiplicity of NLs**: why should the hyperuranus (i.e., the platonic place where real categories exist) be written in English and not in Italian, Chinese, or Tamil?[15] Let me explain with a concrete example. The concept of BLUE can differ substantially if we change the NL in which we are describing it. For instance, in Rumanian BLUE is *albastro*, derived from the Latin *albus*, 'white'. The ancient Romans probably saw the sky colour as a kind of 'dirty white', as the Italian word *celeste* show evidence of. In fact, the English word *blue* is of German origin.[16] Hence, not only is it impossible to a priori decide which categories have the right to enter the hyperuranus, but it is also impossible to decide in which NL they have been formulated. Even if the inventory of phonemes of a given NL is limited and can be reasonably identified, the inventory of a priori semantic categories can not.[17]

On the other hand, the very concept of NL is a (reasonable) abstraction for the continuum variance of idiolects, i.e., the linguistic habits belonging to each person. This fact undermines the very foundations of radical nominalism. If each person has his/her own linguistic convention without any further level of abstraction, it is very difficult to explain how we can (mis)understand one another in a given NL. Things go even worse when we accept the argument of multiplicity of natural languages: translation becomes a priori impossible. Of course, these problems vanish if the aim of linguistics is all and only describing each language system as a monad, i.e., as the unique entity which do not have any perception of the real or even mental world, as Ferdinand de Saussure seemed to believe, according to his famous *Cours*.[18] Following Saussure, thought is inherently shapeless and so there is no pre-linguistic concept. In its extreme form, nominalism transforms itself into **relativism**, a position ascribed to Edward Sapir and Benjamin Whorf: concepts are entirely determined by the NL in use, and therefore looking for universal aspects in linguistics becomes nonsensical.

There is a third way, other than nominalism and realism, namely **conceptualism**. Cognitive linguistics take as a start point the philosophical position of conceptualism: between real-world entities (**referents**) and linguistic entities (**sentences**), there is a set of intermediate entities which re-

sides in our mind (**concepts**). Conceptualism has a number of advantages, and that is why I take this philosophical position for adgrams. First, it solves the problem of entities which have linguistic existence but not a real one, such as the mythical chimera: chimeras simply exist as concepts without having a clear referent in the real world. Second, this position lets people share concepts: where there is an agreement about the referents, i.e., there is a linguistic convention, people can use the same linguistic entity in order to indicate the same concept. Therefore, it makes sense to speak of 'cognitive linguistics' as the discipline which analyses the relation between linguistic entities and concepts.

Adgrams are cognitive in the sense that they are indebted to some results found in cognitive linguistics. Cognitive linguistics has for the last 25 years been an alternative to the Chomskyan approach.[19] Its aim is for a cognitively plausible account of what it means to know a NL, how NLs are acquired, and how they are used. Cognitive linguists are careful to have sensible linguistic data to prove their theories, and that is why I take most examples from corpora of language-in-use, following the principle of linguistic adherence (see previous section).

Unlike Chomsky's hypothesis that language is an innate cognitive faculty – known as the Language Acquisition Device (LAD) – cognitive linguists take as their start point that premise that language is *not* an autonomous cognitive faculty and consequently it is not separated from non-linguistic cognitive abilities. To put it differently, the representation of linguistic knowledge should be similar to the representation of other conceptual knowledge. This does not mean that a unique configuration of cognitive abilities devoted to language does not exist; it means that the language ability requires cognitive components that are shared by other cognitive abilities. In other words, an autonomous level of organisation does not necessarily entail modularity.[20] NLs, cognitive linguists say, are driven by established facts about human cognition, not by the internal logic of the theory. Consequently, grammar is considered as a cognitive ability in *all* its parts – i.e., phonology, morphology, syntax, semantics and pragmatics.

This holistic vision of cognition implies that cognitive linguistics take concepts and models from other disciplines – like psychology – to describe even the smallest, subtle difference between sentences. Using the words of Taylor [2002, 11]:

> the very wording that we choose in order to linguistically encode a situation rests on the manner in which the situation has been mentally construed.

I will refer to this principle as the **grammar as conceptualisation principle**.[21] As I understand the literature, cognitive grammar focuses on the following main areas: (i) categorisation; (ii) mental imagery and construal; (iii) metaphor; (iv) inferencing and automatisation.[22] In the field of categorisation, a place apart is given to **figure-ground organisation**, whose prototype is visual perception.[23] Figure-ground organisation inherits interesting characteristics from perception. In fact, perception is highly related to attention (e.g., you can focus your sight on this very word and, at the same time, focus on the periphery of your vision). Furthermore, figure-ground organisation can be reversed (e.g., you can look at this page as a complexly shaped white figure which obscures a black background). Finally, there are several levels of figure-ground organisation (e.g., if you are reading a paper version of this dissertation, your primary figure is the sequence of black letters while your primary background can be the desk you are sitting at; but there will be also a secondary figure against a wider background of the room where you are sitting in).

While cognitive linguists used the figure-ground organisation principle for semantics, adgrams use it to construct adtrees, i.e., for morphology and syntax, that I treat as two sides of the same coin (see chapter 2 for details). Therefore, the scene that is mentally construed by the speaker is encoded linguistically not only in terms of lexical choices but also by the morphosyntactic resources in charge. Let me explain with a couple of examples:[24]

- (1a.) The farmer shot the rabbit.
- (1b.) The rabbit was shot by the farmer.
- (2a.) The roof slopes gently downwards.
- (2b.) The roof slopes gently upwards.

In example 1 the different figure-ground organisation is due to morphosyntax, while in example 2 it is due to lexical choices. In fact, 1a presents the scene in terms of *the farmer* as the figure, and *the rabbit* as the ground of the event *shot*, while in 1b the scene is reversed: *the rabbit* is the figure, and *by the farmer* is merely a circumstance of the *shot* event, acting as ground. It is important to know at this point that the figure does not always coincide with the syntactic subject. In contrast, the difference presented in example 2 concerns how the speaker perceives the roof: in 2a it is mentally viewed from above, while in 2b it is mentally viewed from below. Adgrams give no cues about these pragmatic and semantic aspects as in example 2, but only about morphosyntactic phenomena like in example 1.

If Chomskyan grammars were attacked by cognitive linguists as 'syntactocentric', analogously adgrams can be rightly attacked as 'morphocentric', i.e., they give a central place to morphology. In fact, I argue that morphosyntax is also a key for semantics. Therefore, adgrams follow a **semasiological perspective** in the language-world approach: it goes from the language model to the world, asking 'For this expression, what kinds of situations can be appropriately designed by it?'[25] Of course, linguistic

Table 1.1: Macro-levels of analysis in adgrams (specimens)

pragmatics		
speech act analysis	**semantics**	
actant analysis	semantic relations	**morphosyntax**
cleft sentences	semantic features	adpositional spaces
	lexical values	adtrees

meaning cannot be reduced to a full compositional model, as it involves not only the **construal**, i.e., the process whereby concepts are structured in a given linguistic expression, but also the **content** evoked. In other terms, a linguistic expression can evoke the same content and nonetheless be different in meaning as it is construed in a different way. What I argue is that the construal is compositional, while content is encyclopædic in scope. Adtrees will give account of the choice of construals in terms of trajector/landmark alignment (see Chapter 2). It is right to say that adgrams are meaningful primarily in terms of construals: semantic structure is an inherent part of adgrams.

I have just used the term 'concept'. With respect to concepts, adgrams borrow another expression from cognitive linguistics. According to cognitive grammar, each sentence forms a **conceptual space**.[26] A conceptual space is formed by linguistic elements which are instances of concepts and their relations. The main point is that understanding is prior to judgements of *semantic* relations – such as synonymy, hyponymy, antonymy – and *pragmatic* relations analysed in terms of speech acts; an analysis which involves implication and presumptions, e.g., which mental states are entailed in the each **actant**. To avoid confusion, I will adopt the term 'actant' from Tesnière, to indicate the semantic roles in each phrase, while the term **participant** will simply indicate the two leaves of a minimal adtree (see Chapter 2 for details).[27] Semantic and pragmatic relations form content, while the conceptual space is construed (see Table 1.1). Each adtree aims to describe a conceptual space: content is included in the lexicon, i.e., in the leaves of the adtree (see Chapter 3 for the treatment of lexicon).

My understanding of the literature in cognitive linguists is that semantics drives morphology and syntax, which are considered merely as parts of it.[28] In my view, the main limitation in the current cognitive linguistics approach is the premature and prejudicial refuse to apply a formal methodology to their linguistic analysis. Most of this literature is not precise enough to be used in a language understanding or generation system. At the same time, cognitive linguists use conceptual spaces as starting points to describe semantics in terms of encyclopædic knowledge, an approach which raises many unsolved problems.[29] In constrast, adgrams are driven by visible

linguistic items – the morphemes – and they strictly follow the conceptual space described by each sentence, following the principle of linguistic adherence. Consequently, semantics is derived from the treatment of morphemes in a strict formal way.

Adpositional grammmars are structural

The main problem with conceptualism is the mapping between references, concepts and linguistic entities. Do linguistic entities reflect some properties of the referents (mild realism) or simply reflect concepts in the mind (mild nominalism)? It seems quite odd to visualise concepts as 'pictures in the head', since it is impossible to depict the prototypical concept of TREE, not to mention the representation of concepts like LOVE or GOOD (that is why the allegories of figurative arts are so interesting). We can only visualise instances of the concepts, not their "real" properties. Therefore, we should consider concepts with a purely functional point of view: a concept is a principle of (flexible) categorisation, so that human beings can draw inferences. Following this line of reasoning, each **morpheme** – i.e., each linguistic element with a visible representation which is both atomic and meaningful – is a concept.[30] In 1b these are the concepts involved: *rabbit*, *farm*, *shot*, *was*, *by*, *the*, *-er*. At a first glance, it can sound strange that a derivative morpheme like *-er* is treated as a concept. Let me explain through some more instances of example 1:

- (1a.) The farmer shot the rabbit.
- (1b.) The rabbit was shot by the farmer.
- (1c.) ?The farm shot the rabbit.
- (1d.) The farmers shot the rabbit.

The four sentences depict four different conceptual spaces. In example 1c, the farm is a very unlikely one, equipped with guns and maybe an A.I.: the model of the world behind such a sentence could be a science-fiction novel! What is important is that there is no prior assumption about the model of the world: it is taken as axiomatic in this dissertation that it is always possible to find an appropriate world of reference if the sentence is morphosyntactically sound, i.e., I am agnostic about world models.[31]. Following Chomsky, if the sentence is morphosyntactically sound, it is well-formed, as in the famous example *colorless green ideas sleep furiously*.[32] In example 1d, there are more farmers shooting at the rabbit. The presence or absence of *-er* show that the scene is construed in a totally different way; analogously, the adding of the morpheme *-s* again changes the meaning of the sentence.

The method I will follow in this dissertation is called the method of **collocation** and it is contrastive: two chains of linguistic entities will be offered, with only one difference, and the corresponding adtree will be given. Collocation is a morphosyntactic phenomenon. The method of collocation is typically structuralist: the syntagmatic axis shows the *intra*-relations, in the earlier examples the relations in 1b between *farm* and *shot*; the paradigmatic axis shows the *inter*-relations, i.e., if a different lexical item can be substituted with another one without further change in the syntagmatic relations. In the earlier example, the substitution:

- (1a) *farmer* → (1d) *farmers*

does not involve the other syntagmatic elements, while the substitution:

- (1a) *farmer* → (1c) *farm*

involves the syntagmatic relation between *farm* and *shot*: *farm* has to change one of its moduli from [-ANIMATE] to [+ANIMATE] – for the notion of 'modulus, see section 3.1. However, note for now that the use of discrete units for semantics does not imply that semantics can be reduced to a truth-conditional correspondence with the world model, of which there is no a priori assumption. After a huge collocation analysis, we could induce that the couple *farmer : farmers* is an example of a typical plural marker. That means that the paradigmatic axis is dependent on the syntagmatic axis. Finally, when precision will be needed, I will speak about **substitution** when collocation is applied on the syntagmatic axis, and about **contrast** when I use contrastive couples so as to let paradigms emerge.

Chapter 2 is devoted to make the concept of adposition clear, while the concept of group will be fully exploited in chapter 3. Finally, let me remind the reader that the first part deals with the model from a linguistic point of view, i.e., the rules that will be presented hereafter are language-independent. For a formalisation of the model, a concrete linguistic instance is needed and the reader is addressed to the second and the third parts of this dissertation.

Notes

[1] For the logic of Leibniz, see Couturat [1901]. Boolean logics was published at first in Boole [1848], while Frege's conceptual system in 1879 [Bynum, 1972]. Giuseppe Peano's *Formulario* was completed in 1908 [Cassina, 1960]. For Kurt Gödel's theorem and the Church-Turing thesis, as well as other foundational papers by notable authors, see Davis [1965].

[2] The *lingua generalis* was rediscovered and published by Couturat [1903].

[3] Esperanto was launched in 1887 while *Latino sine Flexione* in 2003 – see Gobbo [2005a] and Gobbo [2008a] in Boers et al. [2008]. For the forgotten pioneer René de Saussure, see Künzli [2001] in Fiedler and Liu [2001].

[4] The masterpieces of the three authors are: De Saussure [1970], Chomsky [1957], Chomsky [1965], Tesnière [1959].

[5] The model was published for the first time in Esperanto in Pennacchietti [1976], and Pennacchietti is still working on it, extending the analysis to new NLs. Nevertheless, the core of the model is solid so that a formalisation can be implemented.

[6] Most of Pennacchietti's research is about comparative analysis of Semitic NLs [Pennacchietti, 1974, 1978], Pennacchietti [1981a] in Cagni [1981], Pennacchietti [1981b], Pennacchietti [1984] in Cagni [1984], Pennacchietti [2005] in Bodgan et al. [2005], and Pennacchietti [2008], while some papers deal with Esperanto Pennacchietti [1976] and Pennacchietti [2006a] in Wandel [2006] and others with Italian [Pennacchietti, 2006b]. The most theoretical paper is Pennacchietti [1984], concerning mainly nominalisation, a grammaticalisation phenomenon (see Chapter 2). I extensively use the Pennacchietti's work throughout the whole dissertation. For readability, I will not refer to the papers unless they touch a specific part of the model.

[7] For example, Langacker declared that he feels a "considerable empathy" for Tesnière's work, even if it is out of the cognitive grammar approach [Langacker, 1995] in [Dondelinger et al., 1995].

[8] Ades and Steedman [1982] proposed the type-shifting-operation, which lead to CCG. An introduction to CCG can be read in Steedman and Baldridge [2007], to appear in Borsley and Borjars [2007]. For TAGs, see at least Jurafsky and Martin [2000]. For HPSG, there is a dedicated web site at Stanford: http://hpsg.stanford.edu.

[9] The only exception that I have made is the brief introduction of Esperanto as a unique sociolinguistic phenomenon in section 4.3.2.

[10] Essentially, I've chosen this multilingual corpus because it contains 16 NLs plus the quasi-natural language (QNL) Esperanto. I will return to this topic later in the dissertation.

[11] I will this crucial point in detail explain in the opening of chapter 2.

[12] For a discussion, see Taylor [2002, :13–16].

[13] The complex story of the famous *Cours* – collected from the students' lecture notes by colleagues of Saussure – does not permit us to say exactly what Saussure thought ([De Saussure, 1970] for a critical edition). In any case, I refer to the position of Saussureanism, i.e., the position that linguists take strictly following Saussure. Saussureanism is still alive and well in continental Europe among linguists devoted to the studies of dead NLs like Latin or Sanskrit. See Taylor [2002] for a strong critic to Saussureanism from a cognitive linguistic perspective.

[14] The debate raised by Eliza was enormous – see at least Russell and Norvig [2002] for a general introduction to the epistemological problems raised by this program. Still, the original paper [Weizenbaum, 1966] is worth reading.

[15] It is interesting to note that traditionally the answer to this question has been given by means of some external reference believed to be the receptacle of Truth, like for instance the Bible. Consequently, the Book of A Priori concepts should be in Hebrew. See Eco [1993] for a lot of interesting insights.

[16] The problem of colour categorization is a classic example. For an account from the perspective of cognitive linguistics, see chapter 1 of Taylor [1989].

[17] Note that even brilliant minds fell into this fallacy, like Leibniz, who tried to find the laws of thought in analogy to algebra [Eco, 1993]. More recently, Jackendoff [1990] and Pustejovsky [1995] fell into the same trap.

[18] I think that the reason for this sharp position was the desire to protect the newborn discipline of linguistics from the more established disciplines at the time that Saussure gave his lectures.

[19] Cognitive linguistics entered the arena of linguistics officially in 1987, when two important monographs were published. Langacker published the theoretical foundations of cognitive grammar, while Lakoff published a treatise on metaphor (see [Langacker, 1987, Lakoff, 1987]).

[20] The hypothesis that language is not an autonomous cognitive *faculty* but 'only' a cognitive *ability* should not be understood as a denial of linguistic innatism, i.e. the fact that learning languages naturally is an innate ability, as clearly stated for the first time by Chomsky [1957].

[21] This principle is referred to as the second hypothesis by Croft and Cruse [2004, :1–4] and it is shared among most members of the cognitive linguistic community. It was launched as a slogan of the cognitive linguistics movement in Langacker [1987].

[22] 'Cognitive grammar' is the most developed cognitive linguistic framework presented by Langacker [1987]. I have rather simplified the overview offered by Taylor [2002, 1989] and the textbook by Croft and Cruse [2004].

[23] Cognitive linguists take from the traditional psychological studies of the so-called "Gestalt school" in Germany. Some authors are Max Wertheimer, Kurt Koffka, Wolfgang Köler, Rudolf Arnheim.

[24] These examples are taken from Taylor [2002, :11], but my analysis is slightly different.

[25] The complementary perspective is onomasiological, i.e., given a state of the world model, it asks what set of linguistic expressions can appropriately describe it. Onomasiological salience will be used in adgrams in the definition of the basic-level term in lexicon – see chapter 3. For a detailed discussion, see chapter 10 in Taylor [2002].

[26] Cognitive grammar is the most structured framework of analysis within cognitive linguistics, originally proposed by Langacker [1987]

[27] For the semantic relations, see Lyons [1977], who follows Gottlob Frege's distinction between sense and reference, where the sense is the set of relations within the lexical system. For the pragmatic analysis, see at least the essay *How to do things with words* by Austin [Austin, 1961] (see also chapter 4 in Lyons [1995]). The term 'actant' was popularised in literary criticism by Greimas [1966].

[28] Even so, cognitive linguists have to recall the notion of syntax as a discrete module in the treatment of idioms (e.g., fixed expressions, formulaic phrases, clichés, etc.) See chapter 27 of Taylor [2002] about this paradox.

[29] Let us take the sentence *John regretted signing the letter* as an example. There is an expressed participant (John), an object (the letter) and a complex action (to regret signing). These are the linguistic units involved which form the conceptual space, together with the speaker and the hearer as additional participants, who cannot be John. All these 'frame' [Fillmore, 1982, 1985] the experience so to build language understanding. Fillmore builds construals in terms of frames; other authors use different approaches. For instance, Langacker [1987] and Lakoff [1987], two pioneers of cognitive linguistics, describe the relations between construals as 'profiles' against 'bases' (Langacker) or 'domains' (Lakoff). For example, the construal *radius*, which denotes the concept RADIUS, profiles against the base CIRCLE. So, the meaning of a linguistic unit should be expressed in chains of profile+base formulae.

Although this approach seems to be promising at a first glance, it has the great disadvantage of looking for construals out of the linguistic units, instead of within them. In fact, if things seem to be simple with the couple RADIUS/CIRCLE, as they refer to mathematical knowledge, which is highly formalised, they fail to give a convincing account in complex

cases such as as BACHELOR/ADULT UNMARRIED MALE, where the encyclopædic knowledge is far more vague (see the whole discussion of this example in Croft and Cruse [2004, :28–32]).

This is a general problem of frame semantics (Fillmore) as well as of encyclopædic semantics (Langacker, Lakoff): as argued by Searle [1979], frames are going to be infinitely complex, as the background presumptions, *out of the actual linguistic units*, presuppose other presumptions as well, and so on, forming a infinite set that is required to characterise the literal meaning of an utterance in its appropriate use in context. That's why this approach was abandoned in the adpositional grammar theory definition.

[30] In this respect I completely agree with Taylor [2002, 42–44].

[31] It was Searle [1979] who introduced the philosophical debate. For instance, in order to properly understand even the most trivial sentences, like *the bottle is on the table*, we have to refer to a non-propositional encyclopædic knowledge, beliefs and desires that Searle calls the Background (for example, the absence of gravity can change completely the meaning of the sentence). The Background cannot be described in terms of truth-conditional sentences, but only in terms of networks. In this dissertation, I will taken the Background as given. See Searle [1992] for details.

[32] It is noteworthy that Tesnière, in a completely independent context, formulated an example in French to clarify the independence of the *plan structural* from the *plan sémantique*: *Le silence vertébrale indispose la viole licite* – cited in [Langacker, 1995]. In Italian, the metapoetry by Fosco Maraini can be considered similar: *il lonfo non vaterca né gluisce e molto raramente barigatta* – cited in Gobbo [1998].

Chapter 2

Adpositional trees

Adpositions are the tools that NLs use to indicate grammatically the relations between linguistic elements. How can we classify adpositions? Do adpositions form a system? The answer to this question bring us directly to the structure of adtrees, and therefore to the foundation of adgrams. It is well known that adpositions combine particular elements in very specific and determinate ways. Any formal grammar should give a precise specification of the elements connected to one another and how they are connected within a single NL. In other words, adpositions should form a system.

As Mel'čuk convincely shows, the relationship between grammar elements can take one of the following two forms, or a combination of them: tertium non datur, i.e., there is no third possibility.[1]

The first form was started by Noam Chomsky. A constituent syntax groups adjacent words into abstract units until reaching the top, i.e., the sentence node. For instance, if we have three adjacent elements such as *a* (an article), *blue* (an adjective), *chair* (a noun), it will be grouped as *a blue chair*, i.e., as a single Noun Phrase (NP). Chomskyan linguistics treats prepositions as appendices to Noun Phrases: they derive a Preposition Phrase (PP) from the formula: $PP = Prep + NP$. For example, *with* becomes the head of the NP *a blue chair*. The great advantage of constituent-based grammars is their relatively easy description in algorithmic terms.

The second form was started, in modern times, by Lucien Tesnière. In this approach, the relation of *blue* and *chair* is expressed in terms of dependency: *blue* is the dependent, *chair* is the governor because in English adjectives depend on nouns. Tesnerian linguistics picks up word pairs establishing their dependency, and acts recursively until all words are related. The theoretical assumption is that "the syntactic structure of sentences resides in binary asymmetrical relations holding between lexical elements." [Nivre, 2005, 6]. In this family of approaches prepositions are treated as modifiers of nouns, i.e., nouns are governors, prepositions are dependents,

exactly as if prepositions were adjectives. There is no grouping of words, i.e., no "phrase" in Chomskyan terms. Compared to the Chomskyan one, the Tesnerian tradition respects cross-linguistic variety more, right from the beginning. In particular, it is more successful in describing NLs with a rich morphology, e.g., German or the Slavic ones. Generally speaking, dependency grammars were better received in continental Europe than in the Anglo-American world.[2] Nonetheless, one of the problems of dependency grammars, derived from the original Tesnerian work, is exactly the vagueness of the notion of dependency: it is not clear how to correlate the 'deep' relations of dependency with the surface morphological order. In fact, most dependency-based frameworks lack proper formalisation for parsing.[3]

Neither the Chomskyan tradition nor the Tesnerian one has succeeded until now in offering a valid description of this crucial point in any NL grammar. Adgrams are made with some concepts from both traditions, combined in a new and original way. From the Chomskyan tradition, linguistic elements are put together in **groups**, roughly corresponding to the Chomskyan notion of 'phrase'. However, groups will be classified according to the grammar character of their head (see section 3.1 for the notion of 'grammar character'). Another element of originality is the fact that adgrams use the Tesnerian notion of valence *without* the notion of dependency. Linguistic elements will be strictly analysed in pairs, but their relation will be described properly by adpositions.[4] This is possible because the atomic unit of linguistic analysis is not the vague notion of "word" – as in *every* computational grammar I have seen by now – but the most robust notion of morpheme.

The rest of this chapter is devoted to explain how adpositions work in building NL grammars.

2.1 Brøndal 's contribution to adpositional spaces

An interesting proposal to classify prepositions was published by the Danish logician Viggo Brøndal, independently of any school of research in theoretical linguistics, and it is still unknown in the current linguistic debates, as far as I know [Brøndal, 1940]. Pennacchietti's system is a derivative work of Brøndal's. Brøndal considered two fundamental relations of logic in order to classify prepositions: transitivity and symmetry. Transitivity is the first axis that depicts Brøndal's prepositional space. For instance, the English preposition *in* is always transitive, because it is possible to construct a sentence like example 6, which is well-formed, although rather unlikely, while example 7 is completely ungrammatical.

- (6.) There are no more chairs *in* the room *in* the trattoria *in* the town *in*...

- (7.) *They stopped *at* the trattoria *at* the town *at*...

Therefore, the English preposition *in* is transitive, while *at* is intransitive. Symmetry is the other fundamental axis for Brøndal. A preposition is symmetrical when the semantic relation between the two participants is equal.

- (8.) They ate a Sicilian sea bass *with* a bottle of white wine.
- (9.) They ate a Sicilian sea bass *without* (a bottle of white) wine.

The English preposition *with* is symmetrical, as the relation in (8) from the *Sicilian sea bass* to *bottle of white wine* is symmetrical compared to the relation from the *bottle of white wine* to the *Sicilian sea bass*. Conversely, in (9), the case of *without* the relation between the *sea bass* and the *wine* is obviously asymmetrical.[5] The two axes form Brøndal's prepositional

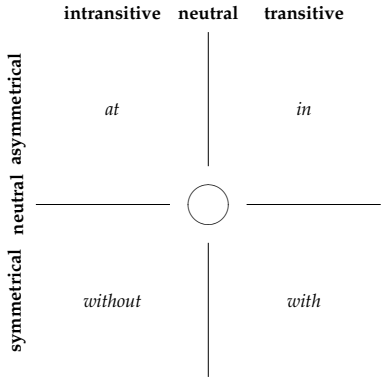


Figure 2.1: Brøndal's prepositional space of English (elaboration)

space; if a preposition shows two different behaviours, it will fall into a neutral box (see Figure 2.1).[6] However, the criteria proposed by Brøndal cannot explain the whole set of prepositions of a given NL, but only the so-called 'proper' prepositions, as admitted by Brøndal himself. Hence, Pennacchietti looked for a different criterion to classify adpositions following Brøndal's methodology. The criterion behind the new system is the dichotomy trajector/landmark.

2.2 The dichotomy trajector/landmark

Cognitive linguists give an interesting insight into prepositions. In fact, a preposition is viewed as an atemporal Gestalt (i.e., a pattern or model organised as a whole). In particular, Langacker [1987] describes a preposition as an atemporal relation between the participants in an utterance. These atemporal relations are asymmetrical. The most salient participant is in the focused position called the **trajector (tr)**, indicated by a triangle (△).[7] The trajector cannot be express without a reference point of observation: this is called the **landmark (lm)**, indicated by a circle (◯). The tr is the primary focal participant, while the lm functions as ground to the trajector.[8] Furthermore, the lm can act as the secondary focal participant recursively, if there is a less salient lm to be referenced to.[9] In spite of this promising premise, most efforts of cognitive linguistics are not concerning formalisation and therefore their observations are beyond the scope of adgrams.[10] Nevertheless, this dichotomy can be used to form a linguistic system suitable for formal and computational treatment, as shown in sections 2.3 and 2.4.

2.3 Applicational adpositions, applicative landmarks

An example scenario will be used to show how NLs represent can represent the same content in different conceptual spaces. The scenario is made by a sequence of contrastive examples with a specific use of adpositions.[11] The conceptual space behind the scenario has three participants: Mr. A, the Book B, and Mr. C. In example 10 the book B is the tr (△), while Mr. A and C are the lm (◯). The verb is given in brackets because it can be omitted.

- (10-en.) △(A book) (is) ◯(*between* two men).
- (10-it.) △(Un libro) (è) ◯({*tra*|*fra*} due uomini).
- (10-tu.) ◯(İki adam) △(arasın*da* bir kitap).

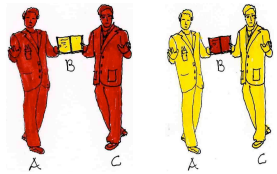


Figure 2.2: The *between* relation (10, left) and the *VO* (11, right).

Not all lms are equal. When an adposition profiles the tr on the lm actively, I will say that the tr is applied on the lm. In other terms, the English *between*, the Italian {*tra*|*fra*} (which are equivalent), the Turkish *-da* are adpositions profiling **application**. By now, in Figures like 2.2, the convention will be that the tr is always **yellow**, like sunlight, while the **applicative lm**, i.e., the lm defined by an applicative adposition, is always **red**, because the relation is active (think of red as an icon of 'hot water').

In example 11 I have exchanged the tr and the lm: A and C are now tr, while B is the lm (see also Figure 2.2, right). Note that this adposition marks the important role between the verb (V) and the object (O) in two different ways: English and Italian sign it by collocation, while Turkish uses a specific morpheme. If an adposition signs the relation by collocation it will be called a **zero adposition**. As the reader can see, I have updated the definition of adposition given in the Introduction: adpositions include prepositions, postpositions, circumpositions and also zero adpositions. Furthermore, adpositions profile relations to any kind of lexemes, not only nouns: for instance, the relation SV has been established within a stative, i.e., 'nominal', group and the verbal group; moreover, the function of some adpositions is to interrelate whole phrases (see the last section 2.10 of this chapter for details).

- (11-en.) △(Two men) hold ◯(ϵ a book).
- (11-it.) △(Due uomini) tengono ◯(ϵ un libro).
- (11-tu.) △(O iki adam) ◯(kitabı) tutuyorlar.

Action *hold* is the baseline where tr and lm can operate. Note that in English and Italian there is a zero adposition, which is part of the group *the book* because it signs its role as a syntactic object O.

The change from example 10 to example 11 implies a change in the nature in the relation. In fact, while the relation stated in example 10 is always *in presentia*, the relation stated by example 11 can be *in absentia*. Consider the following contrastive sentences:

- (11-en.) △(Two men) hold ◯(ϵ a book).
- (12-en.) △(Two men) own ◯(ϵ a book).

Figure 2.2 can apply or not apply to example 12, as the *book* (lm) is owned by the *two men* (tr) even if they do not hold it in hand. In contrast, the adposition *between* can never be used *in absentia*, even if the relation is abstract. For example, in the group *the relationship between Liza and Paul*, the adposition *between* states metaphorically that the relation is present and real between Liza and Paul.

If an applicational adposition should always be applied *in presentia*, like *between*, that application is called **dimensional**; otherwise, that application will be defined as **non-dimensional** (expressions coined by Pennacchietti).[12] I will refer to this step in the theory of adgrams as the **first reflexion**. The first description of the applicative adpositional space of English is represented in Figure 2.3. Applicative dimensional adpositions will be called

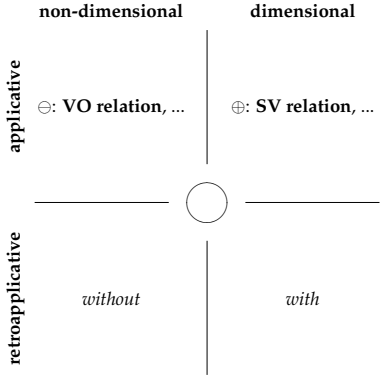


Figure 2.3: A provisional applicative adpositional space of English

Plus (⊕), while applicative non-dimensional adpositions will be called **Minus (⊖)**. These are the lists of the main Plus and Minus prepositions of the English language:

- **Plus**: *in*, *on*, *above*, *over*, *before*, *against*, *about*, *between*, *through*, *throughout*, *among*.
- **Minus**: *like*, *at*, *to*, *at*, *for*, *till*, *until*.

SV relation for Plus and VO relation for Minus are marked in bold, as the prototypical adpositions of the respective subspaces. Furthermore, they do not belong particularly to English. In fact the prototypical Plus and Minus adpositions define the relations of the following language universals: subject (S), verb (V) and object (O).[13]

However, there is a limitation in the current model of adgrams. They do not take into account the distance from the prototypical adpositions, nor do they use neutral boxes, unlike for example Pennacchietti [2006b]. In other words, subspaces are not homogeneous: some adpositions represent better than others their own subspaces. Nonetheless, I treat adspaces not

as radial categories but as binary categories. In other words, in the actual state of adgrams, i.e., the one described in this dissertation, I use the classical approach to categorisation instead of a prototypical one, essentially to preserve elegance and clarity in the formalisation presented in the third part of this dissertation.[14] The dimensional axis is the first one for defining adpositional spaces, and it corresponds roughly to Brøndal's transitivity. Prepositions *with* and *without* are put in Figure 2.3 as in Brøndal's prepositional space of English (Figure 2.1). In order to populate completely the adspace, the second axis is needed. This is the topic of the following section.

2.4 From applications to retroapplications

At a purely theoretical level, NLs can work only with applicative adpositions. For example, the following text is morphosyntactically well-formed:

> Something is unimaginable. Something has happened. Something displeased Europe. Europe is arrogant. Europe thought Africa so poor. Africa would agree to anything. Africa has said no in pride. Pride is rebellious. Africa has said no to the the Economic Partnership Agreements (EPAs). EPAs are a straitjacket. Africa has said no to trade. Trade liberalised completely. Africa has said no to a pact. A pact has manifestations. Manifestations are latest. Manifestations are colonial.

Although well-formed, the text sounds odd: no NL uses only applicative adpositions to profile relations. This is the original article, taken from the English edition of *Le Monde Diplomatique*:[15]

> The unimaginable has happened, to the displeasure of arrogant Europe. Africa, thought to be so poor that it would agree to anything, has said no in rebellious pride. No to the straitjacket of the Economic Partnership Agreements (EPAs), no to the complete liberalisation of trade, no to the latest manifestations of the colonial pact.

Human beings have found strategies to present *known* relations between linguistic entities in a quicker way. For instance, the sentence *Peter has a book* is new information, while *Peter's book* denotes the same information as known. In other words, the English saxon genitive case builds a non-active relation between *Peter* and his *book*.[16] This kind of relations are called **retroapplications** by Pennacchietti: this term is appropriate because concepts come back (i.e., 'retro') from the lm to the tr – they are 'retroapplied'.

In the evolution of a single NL, some patterns of use become standard because of their frequency, so that at last speakers perceive them as a single unit: this process is called **grammaticalisation**. Grammaticalisation is a complex phenomenon. Diachronically, adpositions are formed by grammaticalisation of lexical chains that eventually becomes adpositions. If an

adposition is made of a complex chain of signatures, e.g., some of them put at the left side in the syntagmatic axis and others at the right side, I call them **lexicalised adpositional patterns**, or, for brevity, **patterns**.

Usually, grammaticalisation involves a loss of syllables. For example, we can reconstruct the history of the Italian adposition *riguardo* from the series: *per riguardo a* > *riguardo a* > *riguardo*. Analogously, adpositions like *ciononostante*, *invece*, *nientepopodimeno* are the result of grammaticalisation – they derive respectively from the morphemic series *ciò non ostante*, *in vece*, *niente poco di meno*. The Italian language has even a paradigm formed with *-che* of such adpositions; some examples: *anziché* (*anzi che*), *poiché* (*poi che*), *perché* (*poi che*), *affinché* (*a fin che*), *altroché* (*altro che*) and the archaic ones *epperciocché* (*e per ciò che*), *imperocché* (*in però che*) used mainly in seicento. English examples of analogous grammaticalisation phenomena can be adpositions like *without* (*with out*), *however* (*how ever*) or *nevertheless* (*never the less*). At most, grammaticalisation cuts off every morpheme: these are the zero adpositions.

Retroapplications are signatures of grammaticalisation: their profiled relation is no longer active. For this reason, I will adopt the convention that the **retroapplicative lm** is always **blue**, (think of it as an icon of 'cold water'). I will refer to this step in the theory of adgrams as the **second reflexion**. The second reflexion is applied to the whole axis, i.e., both to Plus and Minus subspaces. Therefore, retroapplicative adpositions can be either dimensional or non-dimensional. Now, it is possible to draw the

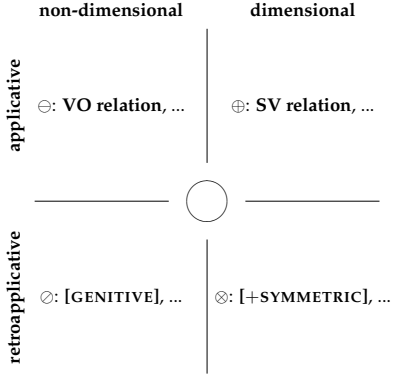


Figure 2.4: The structure of adpositional spaces

whole structure of adpositional spaces, i.e., **adspaces**. Figure 2.4 shows the structure of adspaces. These are the lists of the main Slash and Times prepositions of the English language:[17]

- **Slash**: *without*, *of*, *'s*, *than*, *from*, *off*, *by*.
- **Times**: *with*, *beside*, *near*, *under*, *after*, *behind*.

The prototype of Slash (⊘) is the relation marked in many NLs as the genitive. Slash adpositions are non-dimensional: for example, in *Peter's book* the *book* belong to Peter even if he is not present, i.e., it is a relation *in absentia*. Most adjectival relations fall into the Slash subspace: adjectives, numerals, determiners, articles, some deictic expressions (see chapter 3 for details). The Slash subspace corresponds roughly to Brøndal's intransitive non-dimensional box. For example, some English Slash adpositions are: *off*,[18] *than*, *from*, *without*; they are all intransitive, in Brøndal's terms.

In contrast, Times adpositions (⊗) are dimensional. We do not have a proper prototype, but we can extract the semantic feature of [+SYMMETRY] from Brøndal's taxonomy as the underlying characteristic of Times adpositions: *near*, *with*, *and*, are typical representatives of this category. Let me explain through an example.

- (12-en.) Two men (are) *with* a book.
- (12-it.) Due uomini (sono) *con* un libro.
- (12-tu.) Kitap*lı* iki adam.

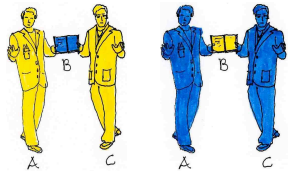


Figure 2.5: The *with* relation in example 12 (left) and 13 (right).

The relation instanced in 12 by *with* (English), *con* (Italian) and *-lı* (Turkish) is equivalent: Mr. A and C are the tr, while the book B is the lm (Figure 2.5, left).[19]. This relation is also symmetrical, in Brøndal's terms.

- (13-en.) A book (is) *with* two men.
- (13-it.) Un libro (è) *con* due uomini.

- (13-tu.) İki adam *ile* bir kitap.

In example 13 trs and lms are inverted: the book B becomes the tr, while Mr. A and C become the lm (Figure 2.5, right). This inversion is impossible with Slash adpositions (**book's Peter*), as well as in the case of Plus (*?two men are between a book*, see 10) and Minus (*?a book holds two men*, see 11).

2.5 Primary and secondary landmarks

Adspaces are like paradigms: they need to be instanced by lexemes in order to form conceptual spaces, i.e., to be put on the syntagmatic axis. From now on, the reader has been introduced to the technical terms, so that I can define collocation as follows:

> **collocation**: *a chain of lexeme pairs whose relation is expressed in terms of tr/lm and morphologically signed by an adposition.*

The triple lexeme-adposition-lexeme forms a **syntagm**. This kind of formalisation was proposed at first by Silvio Ceccato, although he used a different terminology – i.e., Correlational Grammar.[20] Therefore, we can rightly say that collocation is a chain of syntagms. However, syntagms are not at the same hierarchical level. Adtrees are the tool to represent the hierarchy of syntagms. An immediate corollary is that adtrees are formed by at least one syntagm.

Let me explain going directly to our example scenario. I will consider A, B, and C as three different actants: Mr. A is the agent, Mr. B is the beneficiary, while B is the vehicle of the benefaction. Let us see examples 14 and 15, with the following caveat: I give only a partial representation in immediate constituents for clarity; more specifically, in example 14 only the Minus adpositions are represented explicitly, signing the VO and the beneficiary relations, while in example 15 only the non-dimensional retroapplicational adposition {*from*|*da*|*-dan*} is made explicit. All other adpositions are omitted. Conventionally, zero adpositions are signed by the appropriate symbols, while non-zero adpositions show the adpositional morpheme(s) in brackets.

- (14-en.) Mr. A gives ⊖ the book ⊖(to) Mr. C.
- (14-it.) Il Sig. A dà ⊖ il libro ⊖(a)-l Sig. C.
- (14-tu.) Adam ba ska bir adam⊖(-a) kitab⊖(-ı) veriyor.

The conceptual space in example 14 is represented in Figure 2.6, left. Mr. A is the tr, while B and C are two different lms. Mr. C is dotted and not coloured because it is a **secondary lm**, subordinated to B. Unlike primary

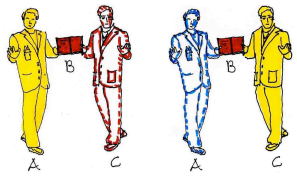


Figure 2.6: Two *give* conceptual spaces: example 14 (left) and 15 (right).

lms, a secondary lm does not need to be mentioned in order to form a sentence well. In fact, while *Mr. A gives the book* is grammatical, **Mr. A gives to C* is not.

- (15-en.) Mr. C receives ⊖ the book ⊘(from) Mr. A.
- (15-it.) Il Sig. C riceve ⊖ il libro ⊘(da)-l Sig. A.
- (15-tu.) İkinci birinci adam⊘(-dan) kitab⊖(-ı) alıyor.

In example 15 the content is conceptualised under a different perspective: C is the tr, B is the primary lm while A is the secondary lm.[21] The verb {*receives*|*riceve*|*alıyor*} distinguish two different lms like {*gives*|*dà*|*veriyor*} in example 14. In fact, in Figure 2.6 (right) A is drawn dotted and not coloured.

2.6 The structure of adtrees

Adtrees are the bread and butter of adgrams. The minimal adtree describes a single syntagm, i.e., a chain lexeme-adposition-lexeme. This order is the conventional one I adopt for adtrees, not necessarily reflected in collocation. It is noteworthy that an adtree is governed by the head, which is the node put "above the root", inherited from Ceccato's intuition, that I call simply the **hook**. The hook carries three pieces of information:

1. the adtree type;
2. the arrow;
3. the adpositional morphemes.

The **adtree type** is one of the subspaces as defined before, i.e., Plus, Minus, Slash, Times. Adtrees are coloured. Colours are not strictly necessary, but they help readers figure out where tr and lm are, with the help of the arrows. The **arrow** is redundant: it is an aid to remind the reader

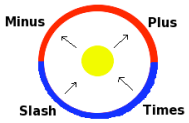


Figure 2.7: The meaning of colours and arrows in adspaces

whether we are dealing with applications, i.e., from the tr to the lm, or with retroapplications, i.e., back from the lm to the tr. The meanings of **adtree colours**, as well as the directions of the arrows, are explained in Figure 2.7. This kind of representation of adspaces has been proposed by Tosco [2006]. **Adpositional morphemes** (adp) bring the adpositional value onto the syntagmatic axis. The head of the adtree is the adpositional morpheme put "above the root". If there are more than one adpositional morphemes, they will be put following their collocation: morphemes attached to participants (i.e., leaves) at the left branch, at first, then the "proper" adpositional morpheme (if any), finally morphemes attached to participants at the right branch. Within the same branch, morphemes adjacent in collocation will be separated by a small dash (-), otherwise they will be separated by a single slash (/); e.g., the German circumposition in *ge-spiel-t* ('play-ed') belongs to a verbal branch and so it will be signed as *ge-/-t*. A **small left triangle** (◁) indicates that the *previous* morpheme(s) have been extracted from the *left* branch of the adtree; vice versa, a **small right triangle** (▷) means that the *following* morpheme(s) have been extracted from the *right* branch. Finally, zero adpositions will be signed by the epsilon (ϵ). The examples that follow in this chapter will help the reader understanding the mechanism.

The taxonomy of adtrees derives from their type. Hence, we will have four types of adtrees: Plus, Minus, Slash, Times. I will always follow this counterclockwise order when presenting an adspace (see again Figure 2.7). Figure 2.8 show the abstract minimal adtrees. If needed, a minimal anonymous adtree will be used, coloured in gray (Figure 2.9). Adtrees can be nested one into another: this structure is isomorphic to the one presented in Ceccato [1961b] (see Figure 2.10; the example means 'fish and meat in (the) fridge since yesterday').

2.6.1 The meaning of the two diagonal axes

Adtrees are built on two diagonal axes, which should not be confused with the axes of collocation, i.e., syntagmatic/paradigmatic. In fact, these axes are built on the four cardinal points: the branch at the left of each adtree is roughly **southwest-northeast (SW→NE)**, while the branch at the right is roughly **southeast-northwest (SE→NW)**.[22] Both axes have their roots

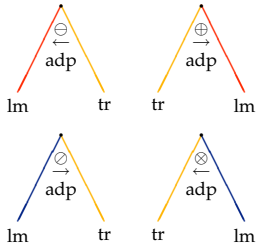


Figure 2.8: The abstract minimal adtree types

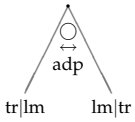


Figure 2.9: The minimal anonymous adtree

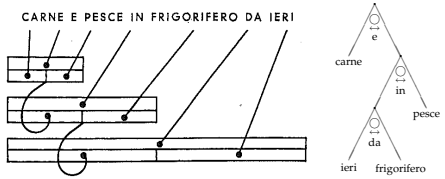


Figure 2.10: The relation between adtrees and Ceccato's translation system

into known information, which is put in the south, while new information is built upon towards the north, according to the order presented in the constructive space.

Conventionally, within a *single* phrase, adtrees are built along the south-east-northwest (SE → NW) axis, while the relations between *different* phrases are rendered along the southwest-northwest (SW → NE) axis. Sometimes adtrees clashes if a branch is not drawn longer; conventionally, I will lengthen the SE→NW axis to avoid clashing *within* a phrase, while the SW→NE axis will be lengthened *between* more phrases. Finally, when I have symmetric junctives, I lengthen *both* adtrees, so to put emphasis on the symmetry.

I use the term **phrase** to indicate a couple of syntagms where the lm is a verb along with its primary tr, i.e., a syntactic subject S, usually espressed by a noun. In order to form a phrase well, the primary tr and lms should be filled.[23] Phrases can be integrated as component structures of **constructions**. Constructions are open-ended until they resolve their composite structure: when this happens, a **sentence** is formed. The conventions of constructions will be given later in section 2.10.

Next sections 2.7, 2.9 and 2.11 will deal with the inner mechanisms of adtrees composed by a single phrase: there, I present the set of conventions to combine adtrees to describe correctly primary, secondary and tertiary lms. The question is: how to determine the inner hierarchy of lms? In order to give the answer, the concept of valence is needed.

2.7 Valence and circumstantials

In cognitive linguistic terms, adpositions profile atemporal relations, while verbs profile processes and hence events, building the scene behind the conceptual space. Therefore, verbs are central to define the hierarchy of lm. Unfortunately, no formal lexical information is provided in cognitive grammar, as authors programmatically do not distinguish between the dictionary and the encyclopædia, i.e., the compositional part of semantics, which belongs to the inner structure of each NL, and the non-compositional part, which pertains the relation between the linguistic system and the world model. In the end, I turned myself again to the Tesnerian tradition.[24]

Tesnière [1959] borrowed the term **valence** from chemistry, which expresses the chemical restraints in combining with other elements. Following the original work by Tesnière, also in adgrams each verbal lexeme has a valence value measured in the number of actants it can accept. Consequently, adgrams can be rightly considered a lexicalised grammar formalism: each lexeme owns a set of values like the grammar character and valence (see chapter 3 for details).

2.7.1 Valence and its actants

Linguistic terminology differs a lot in give account of the number of actants a verb can take: some authors speak of 'intransitive vs. transitive vs. ditransitive' verbs (Taylor), while others speak of 'unaccusative vs. accusative vs. biaccusative'. I will treat verbal forms only on a valence basis. Groups can have valence zero (zerovalent), one (monovalent), two (bivalent) and three (trivalent). Valences cannot be more than three.[25]

Valence is a dimension that should be interpolated by the values of actants, which I present in order of 'trajectivity', from the most prominent, to the less salient, i.e., chain of lms.[26] Actants are the bridge between syntax and pragmatics (see again Table 1.1). Actants should be considered as a **pragmatic layer** appended to syntax, i.e., *after* the building of construal through the mechanism of adtrees.

- **Agent** (N_a): an entity which investigates or which provides the energy to the process, to affect the Patient for the sake of the Beneficiary.
- **Beneficiary or Experiencer** (N_e): an entity which is the locus of the process performed by the Agent.
- **Patient** (N_p): an entity which is affected by the process, generically in terms of a change in state.

N_a,N_e,N_p are the most important actants. For this reason, all nodes representing verb and its actants will be printed in bold. However, there is a further actant I need to explain the theory so far:

- **Instrument** (N_x): an additional entity used by the Agent to perform the event described in the process.

Actants are attached to the most salient tr of a group. A group can have at the most *one* actant (see section 3.5 for details). Actants are all binary categories, e.g. an actant is either an Agent or an Experiencer. The role of actants depends on the valence value.

2.7.2 Zerovalent and false monovalent verbs

Adtrees are Porphyrian trees.[27] The simplest form is zerovalent. Figure 2.11

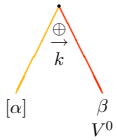


Figure 2.11: The abstract adtree of zerovalent verbs.

shows the abstract adtree. The symbol k indicates the morphemic signature of the adposition, while V^0 indicate a lexeme used verbally, which valence value is zero. I will use greek letters as alpha (α) to indicate lexemes, i.e., morphemes put in leaves, while square brackets indicate arguments that are fixed, i.e., they are idiomatic forms. Verbal leaves are marked with V^x where x is the valence value, while leaves with different fundamental moduli will be marked with their actants. Let me explain through some examples.

In Italian, there is a class of selective zerovalent verbs defined by the semantic feature [+WEATHER], like *piovere* ('to rain'), *nevicare* ('to snow'), while in English they are monovalent:

- (16-it.) Piove.
- (16-en.) It rains.
- (17-it.) Nevica.
- (17-en.) It snows.

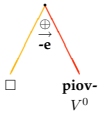


Figure 2.12: The adtree of (16-it) *piove*.

In Figure 2.12 the reader can see a minimal zerovalent adtree. The lm of the syntagm is always indicated by the leaf at the right branch of the given Porphyrian tree, because conventionally it is the most known, or less salient, information, while the tr is put at the left leaf being the most prominent one. This is a general convention that is important for combining minimal adtrees (see later in this chapter). Finally, the head of the syntagm is inherited by the adposition, which gives the relation character to the tree itself.

Even with zerovalent verbs, the prototypical SV relation is actived. This happens because an adtree is *always* Porfirian, i.e., it can never have only "one branch". Let me explain through an example in Italian. The SV relation is marked in example 16 by the verbal morpheme *-e* and in example 17 by its equivalent signature *-a*.[28] The verb in the Italian examples (16) and (17) is put as the lm, while the tr is empty – see the box (□) in Figure 2.12. Non-Italian readers can figure out the scene as in some Japanese paintings, where there is only ground, and no figure emerges. The situation in the English sentence 16 is different. The syntactic subject *it* bears no role in terms of actants, i.e., it is purely syntactic, it has no semantic weight. There is no

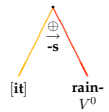


Figure 2.13: The adtree of (16-en) *it rains*.

actant at all, it is a void mask in the scene depicted by the conceptual space. In fact, it is a priori fixed – see square brackets in Figure 2.13. That's why I call **false monovalent** English verbs like *to rain*.

2.7.3 Monovalent verbs

True monovalent groups are quite common; the first valence marks the syntactic subject (S), which is always represented by a stative group whose adposition is Plus. In terms of actants. Figure 2.14 shows the prototypical

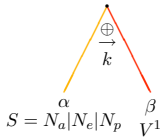


Figure 2.14: The abstract adtree of monovalent verbs

abstract adtree of monovalent verbs.

In monovalent verbs, S can be either Agent, or Experiencer, or finally a Patient. This value is written as a semantic feature in the lexicon (see chapter 3). Let me show some examples.[29]

- (18.) Carl ⊕ {blinked|waved|coughed|meditated|yawned|smiled}.
- (19.) Flowers ⊕ {grew|budded}.
- (20.) The wind ⊕ {fell|rose}.

In all three cases the subject S bears Agenthood, while Patientivity and Experience are not exploited (Figure 2.15). It is important to note that valence

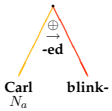


Figure 2.15: The adtree of *Carl blinked* (18.)

and actants are two sides of the same coin, i.e., the vehicle of this kind of

information is the *lexical* information inscribed on the lexemes *to blink*, *to wave*, etc., not the adtree. In other words, the dictionary will dictate which actant value will be put in every single branch of the adtree, as it will be explained in chapter 3.

2.7.4 Bivalent verbs and their actants

Bivalent verbs are another class of verbs. The first valence marks the syntactic subject S, exactly as in the monovalent verbs, while the second valence *usually* marks the syntactic object O as a Minus adposition. I say 'usually' for two reasons. The first one is paying duty to the proviso of a limitation of the current model of adgrams, that is the nominative-accusative schema, that excludes NLs like Basque. This does not mean that adgrams cannot be adapted to Basque: it means that it was not built with ergative-assolutive NLs in mind. Nevertheless, the analysis of actants can give hints to readers interesting in this topic. The second one is the existence of different types of second valences, that I will treat later as a special case study. Figure 2.16 shows the abstract tree of bivalent groups. Let me explain

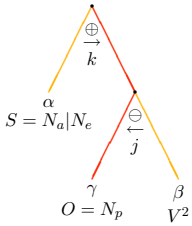


Figure 2.16: The abstract adtree of bivalent verbs.

through some examples (21–24), that depict the same content in different conceptual spaces (not all adpositions are explicited in inline description, for readability, while adtrees are always complete).

- (21.) The ⊕ {phone|bell} rang.
- (22.) Some-body ⊕ rang ⊖(at) the ⊘ {phone|bell}.
- (23.) Liza ⊕(was) called ⊕(at) the ⊘ {phone|bell}.
- (24.) Some-body ⊕ call-ed ⊕(at) the ⊘ {phone|bell}.
- (25.) Some-body ⊕ call-ed ⊖(me) ⊖(at) the ⊘ {phone|bell}.
- (26.) Some-body ⊕ call-ed ⊖(me) ⊗(with) a ⊘ mobile.

- (27.) Some-body ⊕ call-ed ⊖(me) ⊖(at) the ⊘ {phone|bell} ⊗(with) a ⊘ mobile.

The verbal lexeme *rang* in examples 21 and 22 is monovalent, while *called* is bivalent. The adtree of example 21 is analogue to the one shown in Figure 2.15, except of the presence of the article *the*, which is a Minus adposition. What is important to notice is that in examples 22,23 the syntactic subject S {*phone*|*bell*} bears the role of Instrument (N_x), exactly because the lexeme *rang* is bivalent, unlike 21, where the three roles are packed one over the others. More specifically, in example 21 the verbal lexeme is monovalent and therefore {*phone*|*bell*} is the Agent.[30] In fact, unlike monovalent ones, bivalent verbs can show different semantic roles in the phrase: the subject S can be Agent (22, 24, *-body*), as well as Patient (*Liza*, 23).

If the lexeme is bivalent, a collocational analysis, i.e., a parsing, should be performed at first. In example 24 the lexeme *call-* is **saturated** by *me*, i.e., its two valences are fulfilled, being *me* the Patient. In this case, the situation is quite easy to solve: the first valence is the Agent while the second one is the Patient.

2.7.5 The convention for circumstantials

Let us see the adtrees of the example above. Figure 2.17 (left) shows that

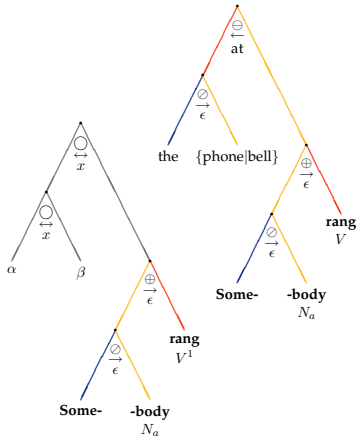


Figure 2.17: The dynamic adtree of *Somebody rang at the* {*phone*|*bell*} (22).

initially example 22 is similar to 18. Now, the phrase is enriched with other elements. What happens if valences are already saturated? We have just seen how the notion of valence gives account of primary and secondary lms. However, NLs have strategies to specify situations beyond the internal arguments of the verb: these are called circumstantials, which acts as **tertiary lms**.[31] Hence, as valences are already saturated, in order to enrich the phrase, the convention of circumstantials will be activated. I said 'convention' as it is an informal description for human beings, not necessarily correspondent to the formal rules I will introduce in the second part, which are machine-readable.

> **Convention for circumstantials**: *if the valences are saturated, the current phrasal adtree is extendend upwards according to the tr branch signed by the adposition: non-dimensional (Minus and Slash) in the SE→NW axis, while dimensional (Plus and Times) in the SW→NE axis; attach non-dimensional circumstantials before dimensional ones.*

Applicative circumstantials are solved before retroapplicative ones because they are more closer to the phrase kernel, i.e., the verb and its valence(s). Conventionally, nodes of circumstantials adtrees are printed in plain, while nodes expressing valence actants are rendered in bold.

In example 22, the adposition *at* is Minus, therefore the adtree will be as shown in Figure 2.17 (right). The branch of the tr in the *at* adtree has been lengthened for readability – along the SE→NW axis, as example 22 is composed by only one single phrase.

2.7.6 Valence and diathesis transformation

The relation between valence and diathesis is a complex topic which was fully exploited by Tesnière [1959], and a full explanation is out of the scope of the present dissertation. I will give only a single example of the transformation from active to passive voice, and how this influences the valence of the verb. The adtree of example 23 is similar to the one of example 22, as *was called* is a passive voice. Note the small right triangle at the left of *-ed*, which indicates that it is attached to the right branch participant *call-*. The diathesis transformation from active to passive decreases the value of valence of *to call* from two to one (Figure 2.18). Unlike 'pure' monovalent verbs, S in phrases governed by *to be called* has the value of Patient (N_p). It is also possible that S has the value of N_e: for instance, in the phrase *Liza was entrusted* the subject S has the value of Beneficiary. All these informations are put in the lexicon (see chapter 3 for details).

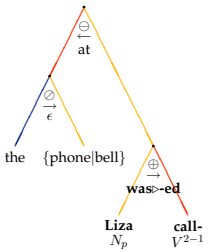


Figure 2.18: The adtree of *Liza was called at the* {*phone*|*bell*} (23).

2.7.7 Valence and circumstantials in comparison

Examples 24,25 are different, as *called* is bivalent. In order to saturate the second valence, the adtree is expanded below, according to the following convention.

> **Convention for Phrasal Adtree Specification (PhAdS)**: *within a single phrase, the adtree is expanded downwards according to the main axis SE→NW: the current leaf node goes to the right, as it is already known, while the novel information goes to the left of the new adtree.*

Figure 2.19 shows how the convention of PhAdS acts. Note that I have extended the SE→NW axis to avoid clashing between leaves. The Minus adtree that governs the syntactic object (*O*) is formed, even if it is not expressed at all. The 'zero' indicates exactly that it was not expressed but it could be. This should not be confused with previous conventional symbols: square brackets indicate that the branch exists but it is blocked, either by a box (□), as already seen in Figure 2.12, or by a fixed morpheme like *it*, as in Figure 2.13.

If the *O* is filled by a morpheme, like in (25), the 'zero' will be replaced by the proper morpheme (Figure 2.19, right). As a side note, PhAdS governs determiners too, and it was already applied to solve the {*phone*|*bell*} group of example 22.

Example 26 clarifies the second part of the convention for circumstantials. In fact, the adpositional value of *with* is Times, and the adtree will be built accordingly (Figure 2.20). The different direction of the Times arrow construes the scene in a very different way compared to example 25. In fact, in example 25 the information provided by the syntagm *at* pertains the actant of the current phrase and it is active, new, therefore it should stay on the left. Vice versa, in example 26 the information provided by *with* is retroapplicational, i.e., non-active and it is a result of nominalisation. More

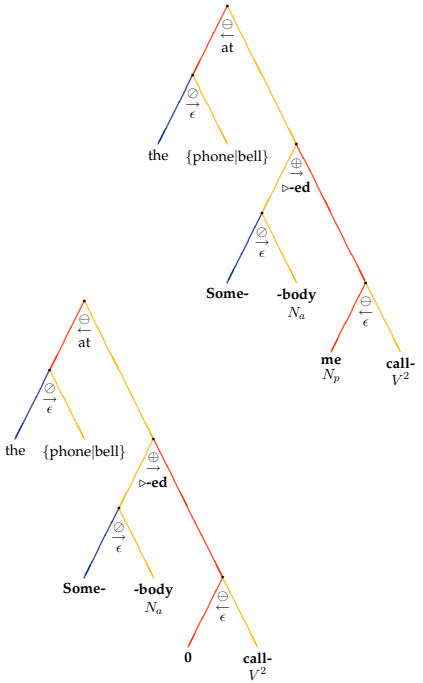


Figure 2.19: The adtree of *Somebody called (me) at the* {*phone*|*bell*} (24,25).

.

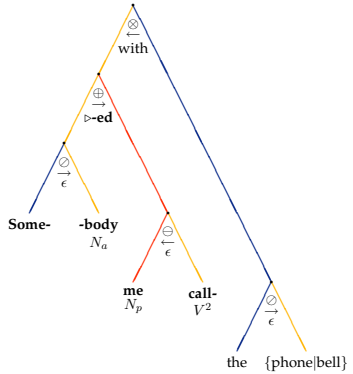


Figure 2.20: The adtree of *Somebody called me with the* {*phone*|*bell*} (26).

specifically, *with the* {*phone*|*bell*} corresponds actively to a whole phrase like 'Someone used a {phone|bell}', and applied back – "retroapplied" – to the main phrase *Somebody called me*, which is considered as a whole tr. Finally, Figure 2.21 shows the whole picture: the two saturated valences are the core of the phrase, i.e., *Someone called be*, while the two circumstantials are parsed following the convention presented above.

2.7.8 Trivalent verbs

Lucien Tesnière was the first grammarian to distinguish clearly between bivalent and trivalent verbs.[32] An verb is trivalent when it marks morphosyntactically the Agent, the Patient and the Beneficiary, i.e., all primary lms. Typically, in NLs with the morphology of case, the third actant N_e is expressed by the dative or by a second accusative: for instance, in Latin *Antonius docet pueros grammaticam* ('Antony teach grammar to the pupils') and in German *das Bush kostete mich einen Taler* ('the book cost me a Taller'). Interestingly, in these examples the Experiencer is expressed by the leftmost accusative, which is closer to the verb and more prominent than the other one.[33] Figure 2.22 shows the abstract trivalent adtree. The symbols means the following: j, j, k are signatures of adpositions; S is the syntactic subject; O is the syntactic object; N_a, N_p, N_e are the three actants; V^3 indicates a trivalent verb. As the third actant N_e is cognitively more prominent than the second actant N_p, it will be put conventionally at a higher position in the adtree.

A proof of trivalence is given by the diathesis, e.g., the transformation

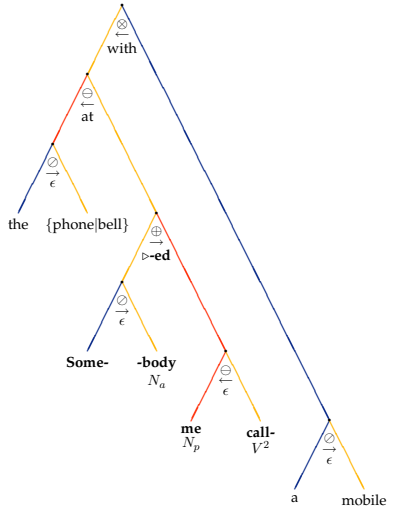


Figure 2.21: The adtree of *Somebody called me at the* {*phone*|*bell*} *with...* (27).

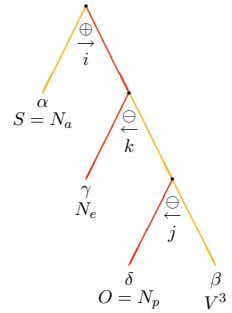


Figure 2.22: The abstract adtree of trivalent verbs.

from the active to the passive voice. In fact, two transformations are possible, as stated in the following example 28, where the actants are put in evidence.[34]

- (28a.) N_a(Al) gives N_p(a book) N_e(to Charles).
- (28b.) N_p(A book) is given N_a(by Al) N_e(to Charles).
- (28c.) N_e(Charles) is given N_p(a book) N_a(by Al).

Example 28a shows the trivalent verb *gives* at the active diathesis (Figure 2.23, middle), while in examples 28b (Figure 2.23, up) and 28c (Figure 2.23, down) the verb *is given* is bivalent. Therefore, the Agent N_a should be expressed by a tertiary lm, i.e., a circumstantial external to verbal valences. Note, that 28c is structurally impossible in NLs like French or Italian.[35] This means that the dictionary entry of the root *give* in the English adgram will have more transfers than the corrispondent roots in Italian or French (see section 3.2 about the concept of 'transfer' and 'transference').

2.8 The most complex phrasal adtree

Figure 2.24 shows the most complex abstract tree having only one verbal node, i.e., consisting of a single phrase. In particular, it shows how to apply the convention for circumstantials with a saturated trivalent verb. In other words, it shows how tertiary lms are ordered – the names of abstract adpositional morphemes follow the order of construction: i, j, k indicate the valences, while l, m, n, o, indicate the tertiary lms. Conventionally, arguments X and Y denote applicative circumstantials while W and Z denote retroapplicative ones – the name of the leaves following the same order of the adpositional morphemes.[36] Of course, there can be more than one circumstantial per type; however, n is usually a small natural number, very rarely more than three, otherwise the phrase would be hardly intelligible, whatever the NL. In this schema nested adtrees are not considered, i.e., adtrees dependent from other adtrees, typically modifiers of stative groups (see chapter 7 for details about this topic).

2.9 Some case studies of adtrees

In this section, I will give some case studies of unusual adtrees. At the end of this section, the reader should have a clear idea how adtrees work in different NLs, even in complex cases.[37]

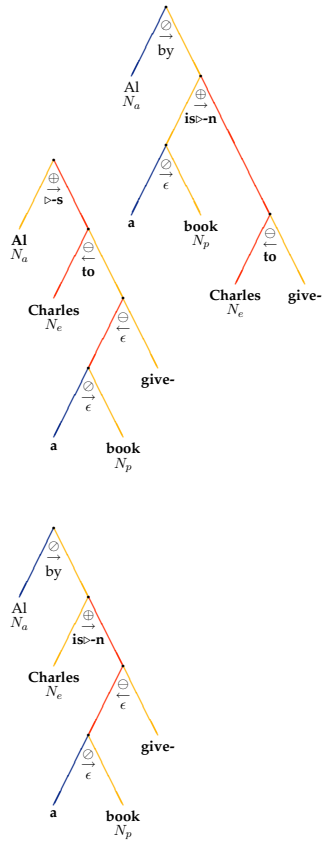


Figure 2.23: The adtrees of the transformation lead by *to give* (28).

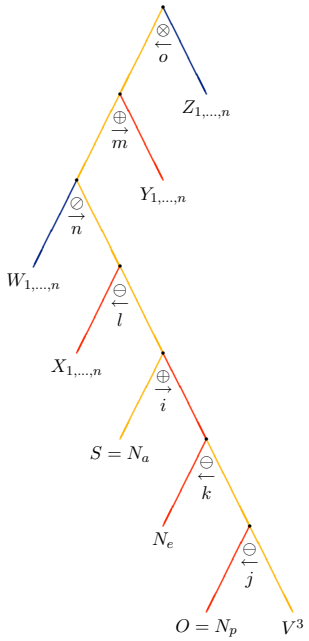


Figure 2.24: The most complex abstract phrasal adtree.

2.9.1 Prepositional bivalent lexemes. False adpositions

In section 2.7.4 I have affirmed that there are bivalent verbal forms whose second valence is not a syntactic object O. In NLs like English or Italian, these second valences are introduced by prepositions. Therefore, adgrams should treat these prepositions *not as normal adpositions but conversely as part of the lexeme*. For instance, an idiomatic form like *fall in love* shows that the lexeme *fall* is bivalent: the English adpositional dictionary should register *to fall* and *to fall in with* as two different verbs: the first one is monovalent while the second one is bivalent. Note that not every preposition governed by *to fall* opens a new valence. Let me explain through an example.

- (29-en.) Liza fell on her knees.
- (29-it.) Lisa è caduta sulle ginocchia.
- (30-en.) Liza fell down at school today.
- (30-it.) Oggi Lisa è caduta a scuola.

In example 29, both English and Italian have a monovalent verb to express the action: {*on her knees*|*sulle ginocchia*} is a circumstance of the falling (Figure 2.25). Vice versa, the Italian sentence 30 does not need a preposi-

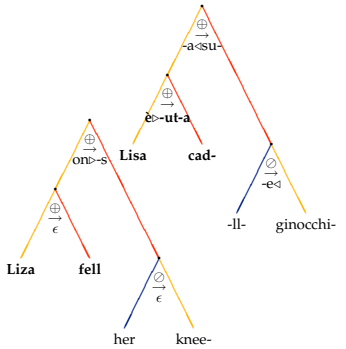


Figure 2.25: The English and Italian adtrees of *Liza fell [...] knees* (29).

tion, while English uses the form *to fall down*. In this last case, *down* is a fixed primary lm, i.e., it works like a second valence that is opened and

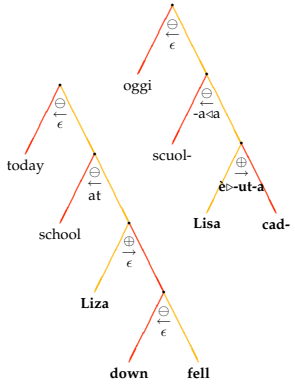


Figure 2.26: The English and Italian adtrees of *Liza fell [...] today* (30).

closed by the morpheme *down* itself. Figure 2.26 shows that example 30 is a purely applicative sentence: a couple of circumstantials determine the boundaries of the falling of Liza. Note an intrinsic limit of adgrams: the English example 30 could also be *Today Liza fell down at school* (30a-en); for the purposes of this dissertation, there is no way to distinguish these two English conceptual spaces through adtrees. It is possible that this kind of nuances are not needed at all, as it is hardly even for a native speaker to distinguish semantics of English examples 30 and 30a. This limitation is even more important in NLs with more freedom in collocation, like Italian: *Oggi a scuola Lisa è caduta* (30a) and *Lisa è caduta a scuola oggi* (30b) are substantially equivalent, in my perception, being a native speaker of Italian. However, in this dissertation, adgrams do not distinguish these constructions of the same conceptual space. There are many form in English that are analogous to example 30, as for instance *to fall off*: the important fact is that they are fixed, i.e., the Minus adtree of *down* in example 30 cannot be extended further. In contrast, sometimes English verbal forms introduce a full second valence, like *to fall in*. Let me explain through a contrastive example:

- (31-en.) Carl falls in(to) the manhole cover.
- (31-it.) Carlo cade nel tombino.
- (32-en.) Carl smokes in the manhole cover.

- (32-it.) Carlo fuma nel tombino.

The English and Italian adtrees 31 are quite similar: in English the only difference is an additional application of the convention for PhAdS to solve *manhole cover* (Figure 2.27). It is noteworthy that the preposition {*in*|*ne-*

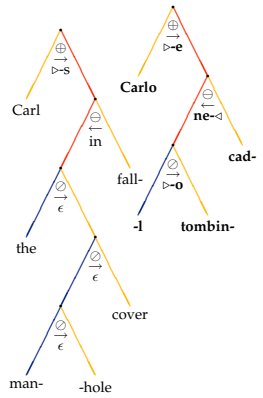


Figure 2.27: The English and Italian adtrees of *Carl falls [...] cover* (31).

} introduces a second valence, hence it takes a Minus value, even if the adposition should take a *Plus* value. In other words, in 31 {*in*|*ne-*} is a **false adposition**. Hence, there is no circumstantial and all nodes are printed in plain. Now, let us take a look at Figure 2.28. Like in example 24, here the second valence is empty: it is not important what Carl is smoking (a cigarette? a pipe? a joint?). In fact, in example 32 the concept depicted by the conceptual space is about Carl who may be hiding for his smoking, therefore *the manhole cover* is cognitively a circumstantial.

2.9.2 To be or not to be? Stative and verbless phrases

In NLs like English or Italian, there are phrases that predicate a description or identification of the subject, usually called 'predicate adjectives' like the following ones:

- (33.) Liza is blonde.

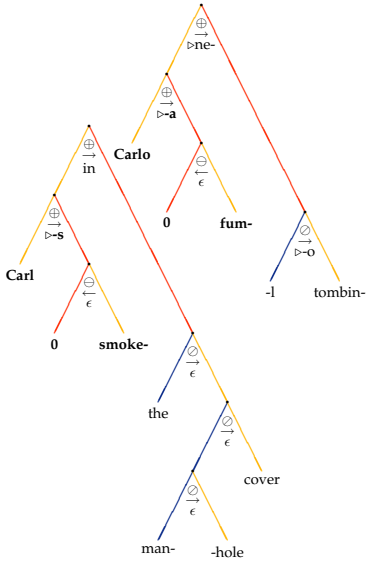


Figure 2.28: The English and Italian adtrees of *Carl smokes in [...] cover* (32).

Analogous constructions are known as 'predicate noun' (in: *Liza became a mathematician*) or 'subject complements' (in: *Liza is mathematician*). I call this collectively **stative phrases**, because they share the same structure in term of adtrees.[38] In fact, in this kind of phrases the verbal forms like *is* or *became* are *bivalent* instead of monovalent, as seen until now. Incidentally, this means that *to be* form a whole family of adpositional morphemes when used as an auxiliary verb *and* a family of lexemes whan used as a 'proper' verb.

This happens exactly for the presence of the predicate noun/adjective or subject complement. The additional valence is treated as a second valence: semantically, it is 'as is' Liza is an Agent and the mathematician is the Patient. The adtree will be construed accordingly – see Figure 2.29.

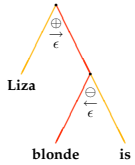


Figure 2.29: The adtree of *Liza is blonde* (33).

Sometimes stative phrases can be a bit harder to be identified. Let me explain through an example.

- (34.) The sky above us.

In example 34 there is an implicit verb *is* signed by the box (□). Note that in this case the implicit verb is not auxiliary, and therefore it takes an additional valence: like in example 33, the preposition *above* is a false adposition – see Figure 2.30. What if there is no verbal form at all? Let us see an example in Chinese Mandarin.[39]

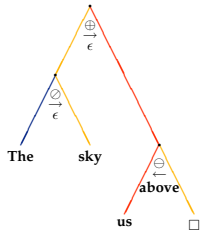


Figure 2.30: The adtree of *The sky above us* (34).

- (35.) shū hěn guì ('books are very expensive').

Example 35 is built morphologically as follows:

shū	hěn	guì
book	*very*	*expensive*

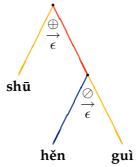


Figure 2.31: The adtree of the Chinese sentence *shū hěn guì* (35).

The lexeme *guì* is monovalentially verbified by collocation, while *hěn* modifies the verbified adjective, as the Slash adposition attests. I call this kind of phrases **verbified phrases**. In conclusion, two distinct subcases of verbless phrases can be found: stative phrases and verbified phrases.

2.9.3 Mirroring

NLs with a relatively high degree of freedom in collocation put interesting problems in complex stative groups, i.e., groups head by a noun with adjectives, numerals and determiniers nested in. In fact, the position of adjective in respect of nouns can mark the situation at a pragmatic level, in terms of knowledge and beliefs of the actants involved. The English language shows a low degree of freedom in collocation, and therefore it uses other strategies, such for instance cleft sentences, so I will explain this phenomenon through some examples in Italian, a NL which shows a higher degree of freedom in collocation than English.

The Italian language distinguish two strategies very clearly with a different collocation of adjectives in respect of noun. Let us draw mentally the following situation. In example 36 there are two Agent, Liza and Bea, who are looking to handsome and not so handsome young men (the Patients).

- Lisa: Quale ti piace? ('Which one do you like?')
- (36.) Bea: Quello alto. ('The tall one')

The sentence 36 is a shortcut of *Quello (che è) alto* (lit. 'the one who is tall'), which depicts two concepts: there is a young man and he is taller than the other. Hence, we have *two* verbless phrases which have lost any lexeme save one. The lexeme *alto* ('tall') is verbified by collocation, i.e., put at the

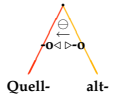


Figure 2.32: The adtree of the Italian sentence *Quello alto* (36).

very right of the phrase, like verbs. That's why Figure 2.32 put it on the right. Normally, the SV relation is expressed by the Plus relation but here the tr is *alto* while *quello* is the lm: the novel information is the fact that the guy is tall. For this reason, the Plus relation is mirrored and becomes a Minus relation. I call this phenomenon **mirroring**, because the verbified lexeme stays on the right but nonetheless it is the attention focus, i.e., the tr.[40]

Let us see a more complex example, which is constrastive. Suppose that two guys, Al and Paul, are talking about common friend.

- Al: Di chi parli? ('Who are you referring to?')
- Paul: (37.) (io parlo) del mio vecchio amico. ('(to) my old friend')
- Paul: (38.) (io parlo) del mio amico vecchio. ('lit. (to) my friend (become) old').

In example 37 Paul's friend is specified by *mio* (my) and *vecchio* (old) as a single piece of information – in fact adtrees apply the convention for PhAdS recursively – while in example 38 the novel information is that he is became *vecchio*: in this lexeme a cleft sentence like *quello che è diventato vecchio* ('the one who became old') is summarised. Therefore, mirroring is applied (see Figure 2.33 contrastively; in brackets the unexpressed verbal form and the syntactic subject).

2.9.4 Presentatives as mirrors. Overvalence

Mirroring is not a prerogative of stative groups. The following example 39 is in Italian:

- (39.) {qui|ora|domani} piove
- (approx. '{here|now|tomorrow} it{'s raining|will be raining}').

According to Figure 2.12, seen previously, which shown the adtree of *piove*, the additional argument {*qui*|*ora*|*domani*} is a circumstantial and the adtree should be construed accordingly (Figure 2.34, left). But, for the effect of mirroring, valence value is augmented by one and the Plus tree is mirrored to a Minus tree (Figure 2.34, right). This augmentation caused by mirroring is called **overvalence**. More examples in other NLs of mirroring are offered, where the adtrees are just like example 39:

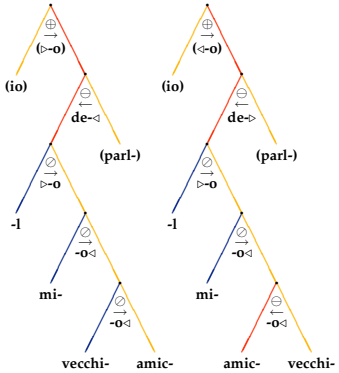


Figure 2.33: The adtrees of *Del mio* {*vecchio amico*|*amico vecchio*} (37,38).

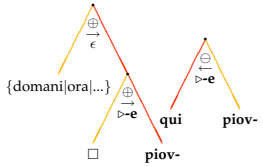


Figure 2.34: The adtree of {*qui*|*ora*|*domani*} *piove* (39).

- (40-la.) ecce homo. ('here the man')
- (40-en.) now you, then him.
- (40-it.) ecco-lo: è arrivato un bastimento carico di...

In the Italian example 40 we have two subsequent presentatives: the lms are *ecce-* (approximately, 'here') and *è arrivato* ('it's come'), while the trs are *-lo* ('him') and *un bastimento carico di...* ('a ship loaded with...'). Note that both English and Italian put the trs in "fossil cases", i.e., cases that are no longer productive. In fact, English uses dative-accusative *him*, while Italian uses the pronominal accusative *-lo*. As a proof, both **then he* and **ecco egli* are ungrammatical.

Mirroring and overvalence solve apperently 'strange' sentences which otherwise could not exist, according to the conventions described until now. Let me explain furtherly through an additional couple of examples:

- (41-en.) it's raining cats and dogs.
- (41-it.) piovono sassi. (lit. 'it's raining stones')

Let us start from Italian. The collocation shows that *sassi* is not *S* of *piovono*: a proof is that **sassi piovono* is ungrammatical. Hence, there are two concepts here: the first concept is *qualcosa che piove* (41a-it) 'something which is raining', that eventually becomes *piovono*, while the second concept is *sassi sono* (41b-it) 'this something raining is stones'. The first component (41a) gives the final lm, while the second one (41b) forms the final tr. Figure

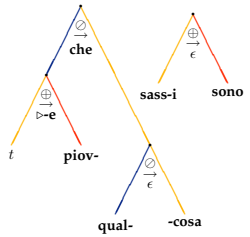


Figure 2.35: The adtrees of *qualcosa che piove* (41a) and *sassi sono* (41b).

2.35 shows respectively the two components. In the left one (41a), there is a node *t* meaning 'trace', which marks the use of the Italian lexeme *che*, which introduces a relative clause. In adgrams, relative clauses are like big adjectives, with the important feature that they solve a valence.[41] Finally, note that the node *piove* in the first part (41a) is a (red) lm, while the node *sassi*

in the second one (41b) is a tr: this fact is respected by the final adtree (41). Figure 2.36 puts together the two components, while Figure 2.37 shows the

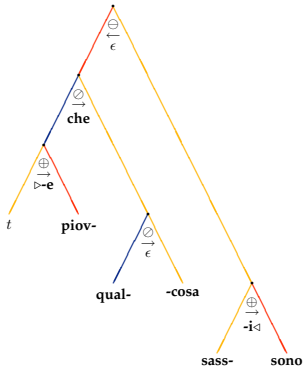


Figure 2.36: The adtree of *qualcosa che piove sono sassi* (41ab).

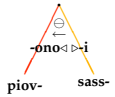


Figure 2.37: The adtree of *piovono sassi* (41).

final, simplified adtree. Note the transformation *piov-e* (singular) > *piov-ono* (plural): agreement is needed, as Italian is a highly inflective NL.

The English example 41 is similar. The first component *it's raining* is the lm (Figure 2.38, left), while the second one *cats and dogs* is the tr – Figure 2.38, right; *and* is symmetrical, hence it is Times.[42] Here there is mirroring *without* overvalence: the behaviour of the tr is like the one of circumstantials, i.e., it is attached above the current adtree (in fact, it cannot be attached anywhere else, because there is only the verbal node *raining* which is semantically non-empty). For mirroring, the Plus node becomes a Minus node: the two syntagms *it's raining* and *cats and dogs* are two different phrases, so they are attached in the way depicted by Figure 2.39. It is noteworthy that morphosyntactic structures like the ones presented in example 41 are often freezed, i.e., they are idiomatic expressions, so they will be perceived as a single, unique block. I will come back to this topic later.

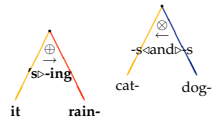


Figure 2.38: The adtree of *it's raining* and *cats and dogs* (41ab).

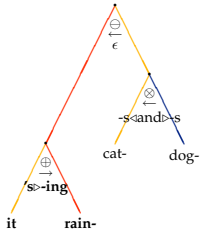


Figure 2.39: The adtree of *it's raining cats and dogs* (41).

2.10 Phrasal adpositions or junctives

When I have put together the two parts of the previous Italian example (41ab) adpositions have been used in a way I did not present yet, i.e., the connector of two different phrases *qualcosa che piove* and *sono sassi*. In fact, until example 40 the reader have been shown only simple sentences, i.e., sentences with only one phrase. What if there are more than one verb in a single sentence? Most language-in-use is made of this kind of sentences, which are usually called **complex sentences**.

In adgrams, the mechanism of adtree functions equally well for the **basic level**, i.e., the structure of the single adtree, and for the **phrasal level**, i.e., how adtrees are joined together. Some adpositions are devoted only to adtrees junction; these are called **phrasal adpositions or junctives**. When I will need to distinguish phrasal adposition to 'ordinary' ones, I will talk about "basic-level adpositions". Some phrasal adpositions in English are {*but*|*however*|*still*|*nevertheless*}, which are all Slash, as the second participant is more relevant than the first one.[43] Other examples of phrasal adpositions are *how*, {*while*|*whilst*}, and the *wh-* series.[44] Note that there is a class of adpositions that can be applied at both levels: two good examples are *and* and *that*.[45] Let me explain through a series of examples number 42, all concerning the same situation.

- (42a-en.) Al knows that book.
- (42a-it.) Alfredo conosce quel libro.
- (42b-en.) Al knows that Carl is studying.
- (42b-it.) Alfredo sa che Carlo sta studiando.

The *that* in example 42a is simply an adjective of *book*, i.e., an element of the lexicon put in a leaf, as clearly shown by Figure 2.40. The English and

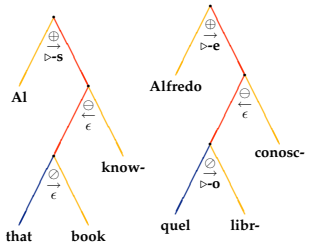


Figure 2.40: The English and Italian adtrees of *Al knows that book* (42a).

Italian adtrees of examples 42a and 42b are perfectly symmetrical. However the English example 42b shows a completely different *that* compared to the one in example 42a: that *that* is a phrasal adposition. In fact, in other NLs, such as Italian, two different morphemes play those roles: *quel : che* – compare Figures 2.40,2.41. More specifically, the second *that* introduces

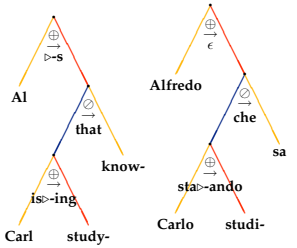


Figure 2.41: The adtree of *Al knows that Carl is studying* (42b).

a whole phrase which is a syntactic object O, so it is expressed as the second valence argument. Finally, note that the phrasal adposition *that : che* introduces the adtree that saturates the second valence of *knows : conosce*.

Phrasal adpositions can be placed in the adspace according to the dichomy tr/lm. If I need to refer only to the placement of phrasal adposition in the adspace, I will talk about **phrasal adspace**. As phrasal adpositions are the result of grammaticalisation, they are all retroapplicative (as explained in section 2.4), therefore phrasal adspace will be only discriminated by the dimensional axis. More specifically, phrasal adpositions can be either Slash or Times. Let me explain through a multilingual example (only trs are signed, through the big triangle).[46]

- (43a-en.) △(Al can pay), because he is rich.
- (43a-fr.) △(Alfred peut payer), parce que il est riche.
- (43a-it.) △(Alfredo può pagare), {perché|poiché} è ricco.
- (43b-en.) Al is rich, {hence|therefore} △(he can pay).
- (43b-fr.) Al peut payer, donc △(il est riche).
- (43b-it.) Alfredo è ricco, dunque △(può pagare).

French *parce que*, Italian *poiché*, *perché*, or English *because* mark the first group as the tr and the second one as the lm, while Italian *dunque*, *pertanto*, French *donc* and English *hence*, *therefore* work in the opposite way: the first element is lm, while the second one is tr. I will not give a complete taxonomy of French, English, or Italian phrasal adspace because it is out of scope of this dissertation. However, it is important to distinguish the basic and phrasal levels of adpositions since now.

In the sequel of this chapter, I will delve into two important phenomena related to phrasal adpositions: splitting and zero junctives.

2.10.1 Splitting

A place apart is given to the phenomenon of splitting, brought by some symmetric adpositions – therefore, they are Times – which can be applied both at basic and phrasal levels. In English, the most important are *(both)... and* and *((n)either)... (n)or*. They are pragmatically very important, because they permit to have more than one actants in the same role at the same time, without valence transformation. Let me explain expanding the previous example:

- (42c.) Al *and* Carl study in the library.
- (42d.) Al studies in the library *and* Carl plays home.

- (42e.) Al studies in the library *and* is sad.

In example 42c Al and Carl are two first actants, i.e., two Agents, within a single verbal form *study*: the adposition is applied at the basic level of the first valence S, while the second one is zero; at the top there is an applicative dimensional circumstantial (Figure 2.42). In contrast, the same

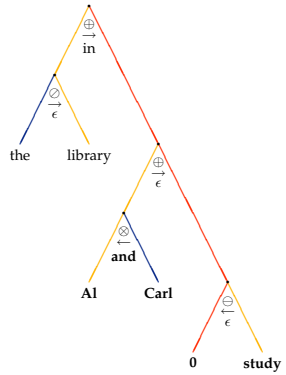


Figure 2.42: The adtree of *Al and Carl study in the library* (42c).

adposition in example 42d connects the two phrases, hence it is applied at a phrasal level (Figure 2.43). Finally, in example 42e Al is a single Agent for two different phrases, and the presence of *and* permit the use of the trace t (Figure 2.44). I call this kind of adpositions **split adpositions** exactly because they let the semantic role be split into two different referents – or more, if applied recursively.[47]

2.10.2 Zero junctives

As previously explained for the basic level, some adpositions do not have a morphemic realisation – they are called zero adpositions. Similarly, there are some junctives with no morphemic realisation. Let me use an example from the classics.

- (44.) vēnī ⊗ vīdī ⊗ vīcī

Example 44 is a complex sentence meaning 'I came, I saw, I conquer', pronunciated by Julius Caesar after a great victory, according to historians Plutarch and Suetonius. The three words are selective verbs, so that each one forms a phrase (Figure 2.45). The word-phrases are simply juxtaposed:

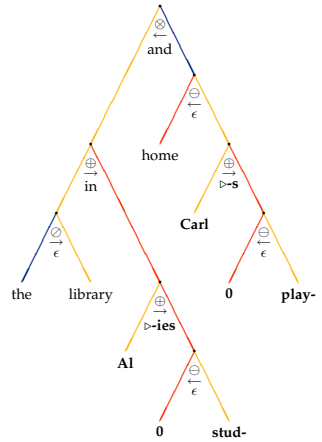


Figure 2.43: The adtree of *Al studies in the library and...* (42d).

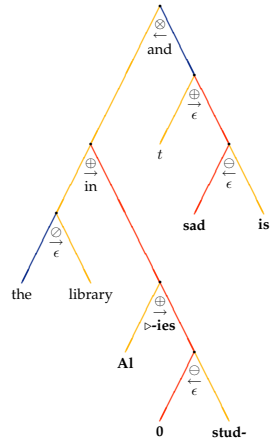


Figure 2.44: The adtree of *Al studies [...] and is sad* (42e).

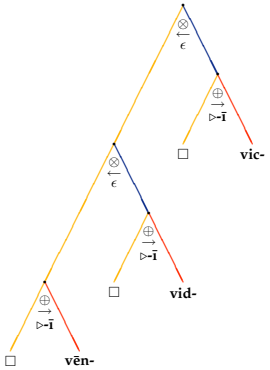


Figure 2.45: The adtree of *vēnī vīdī vīcī* (44).

in Latin the syntactic subject S is put before the syntactic verb V, so the syntactic subject S will be the tr and V the lm. The adposition *-ī* bring different semantic features: person (first), number (singular), tense (past as time, completeness as aspect), so the tr is not expressed. Note that the way in which adtrees are built rightly put in evidence that the last element in juxtaposition is more salient than the others, as the SW→NE axis starts from the known and grows with the new: this textual strategy is known in rhetorics as 'climax'.

Unlike basic-level adpositions, zero junctives can also be marks of punctuation like commas, semicolon, etc., at least in NLs written in some Roman alphabet. Let me show an example in Italian.

- (45.) Si scatenò un finimondo: urla, insulti, scherni, ingiurie.

Example 45 means 'Bedlam broke loose, (and) shouts, insults, mockeries, affronts'.[48] Unlike previous example 44, in example 45 (Figure 2.46) there is a single phrase, so the adtree is built on the SE→NW axis. This is a presentative form, so the adtree is mirrored, and tr is on the right, as explained before, governed by the highest Minus.

The inner Minus of the verbal group *Si scatenò* solves the second valence: this Italian verb is reflexive. **Verbal reflexion** is a verbal phenomenon falling under the rubric of the diathesis paradigm. Adgrams represent it as follows: a bivalent verbal form solves its second valence with a fixed morpheme that acts as a trace of the first valence. In the case of Italian, this fixed morpheme is the proclitic pronoun *Si*. The first valence is expressed by *un finimondo*: all the rest is circumstantial.

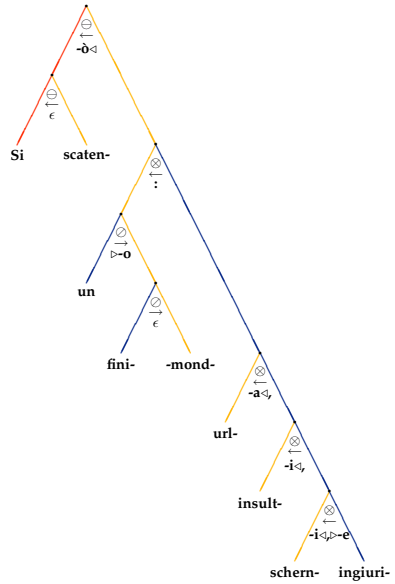


Figure 2.46: The adpositional tree of *Si scatenò un finimondo...* (45).

2.10.3 Phrasal adpositions in different natural languages

The presence of phrasal adpositions is a language universal, but different NLs use different adspatial strategies to construe conceptual spaces and hence render the same content. Let me explain through a multilingual example.[49]

- (46-en.) The library *where* I often study literature is far away.
- (46-it.) La biblioteca *dove* studio spesso letteratura è molto lontana.
- (46-zh.) wǒ jīng chāng xuéxí wén xué. *de* túshū guǎn lí zhér hěn yuǎn.
- (46-jp.) watashi ga bun gaku wo benkyoo *suru* toshokan wa tooi desu.

English and Italian adtrees are symmetrical, save the fact that Italian can omit the syntactic subject S, as the verbal final *-o* marks the person (first) and the number (singular), as Romance languages often do (Figure 2.47). In order to see a different strategy, we should turn to NLs out of the Indo-European family. In Chinese, there is a pro forma question marker which can govern every type of groups, rendered morphologically as *de*.[50] The two main phrases are *wǒ xué xí* ('I study literature') and *túshū guǎn* ('the library is away'). At the intra-phrasal level, note that *zhér lí* 'far' lit. is 'from here', so it should be considered as a circumstantial, unlike Indo-European languages like English or Italian. At the inter-phrasal level, there is a specialised Slash adposition to express the relation without touching the collocation of the components: no overvalence or other strategy is needed (see Figure 2.48, down). Japanese has a phrasal adposition similar to Chinese, *suru* (see Figure 2.48, up). Adpositions *ga* and *wa* indicate the SV relation, but they are not synonyms: for the purposes of this dissertation, it is enough to know that they are both Plus.[51] There is no use of circumstantials in this Japanese sentence.

2.11 A cross-lingual example

NLs differ in organising conceptual spaces. In order to understand well how adpositional spaces influence the way people organise content, let us see a cross-lingual example in parallel. More specifically, this will be given in English, Italian and Turkish.[52] The conceptual space is populated by three actants:

- Actant A: a man, the Agent (N_a);
- Actant B: an unknown book, the Patient (N_p);
- Actant C: man's hands, the Instrument (N_x);

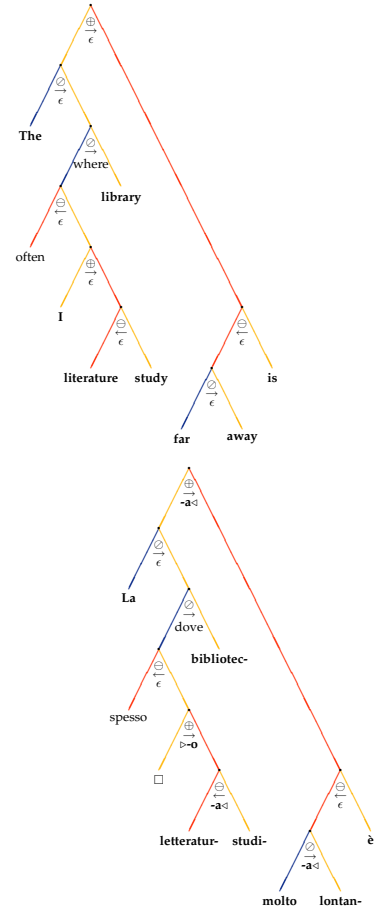


Figure 2.47: The English and Italian adtrees of *The library where I...* (46).

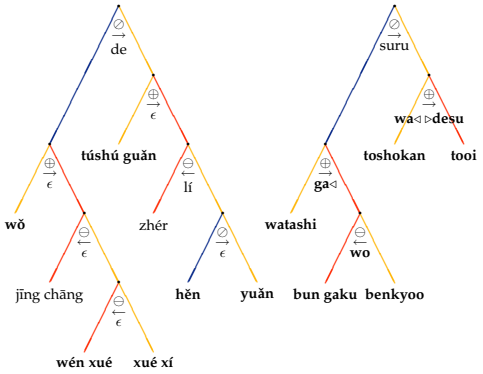


Figure 2.48: The Chinese and Japanese adtrees of *wǒ jīng*... (46).

Table 2.1: Pragmatic analysis of the cross-lingual example

Actants	**English**	**Italian**	**Turkish**
Agent (N_a)	*man*	*uom-*	*adam-*
Patient (N_p)	*book*	*libr-*	*kitab-*
Action (V)	*read-*	*legg-*	*oku-,...*

- Action: reading.

Table 2.1 show some lexemes for the actants involved in the examples I present in the following sections.

2.11.1 A patientive trajector with two different landmarks

Figure 2.49: The Patient as tr with a red lm and a blue lm (example 47).

The conceptual space 47 is shown in Figure 2.49. B is the tr, C is a red lm, i.e., introduced by an applicative adposition, while A is signed retroapplicatively and non-dimensionally. The in-line examples evidentiate the most interesting adpositions, while adtrees show the whole structure of the phrase, as in previous examples.

- (47-en.) A book is ⊕(in) the hand-s ⊘(of) a man.
- (47-it.) Un libr-o è ⊕(ne-)ll-e man-i ⊘(di) un uom-o.
- (47-tu.) Kitap adam⊘(-ın) el-in⊕(-de-)dir.

Figure 2.50 shows that the English and Italian adtrees are identical – save for the leaves. In contrast, the Turkish adtree shown in Figure 2.51 is different. This is the main adspace of Turkish (the brackets are put according to the vowel harmony rules of Turkish):[53]

- **Plus**: *-de*.
- **Minus**: *gibi*, *için*, *kadar*, *üzere*, *-(y)e*, *-(y)i*, *-ce*.
- **Slash**: *-den*, *-(n)in*.
- **Times**: *-siz*, *ile*, *boyunca*.

Example 47 is built morphologically as follows:

kitap	adam-	-in	el-	-in-	-de-	-dir
book	*man*	*of*	*hand*	*his*	*in*	*to be*

The verb *-dir* ('to be') is bivalent, while the main noun *el-* is specified twice by *Adam-* and *-in*. The adposition *-de* is a false one, so to govern the second valence. The rest of the adtree is analogous to the English and Italian ones.

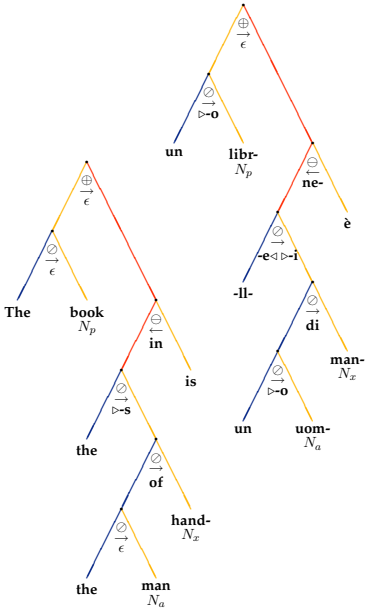


Figure 2.50: The English and Italian adtrees of *The book is in...* (47).

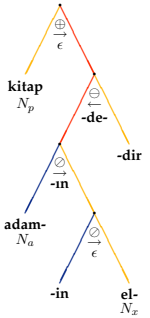


Figure 2.51: The adtree of *Kitap adamın elindedir* (47).

2.11.2 A patientive trajector with two similar landmarks

Figure 2.52: The Patient as the trajector and two red landmarks (48).

Now, let us see the conceptual space of example 48 drawn in Figure 2.52: the book B is still the tr, while the Agent and the Instrument are both red lms.

- (48-it.) Un libr-o è ⊕(in) man-o ⊕(a) un uom-o.

It seems that only Italian can construe this mental space, as far as I know.[54]

2.11.3 Red vs. blue landmarks

In Figure 2.54, the Agent A becomes the tr, while the Patient B is depicted retroapplicatively but in the same spatiotemporal situation, i.e., dimensionally. There is no mention of the Instrument (more precisely, it is part of the Agent). How is this conceptual space expressed in English, Italian and Turkish?

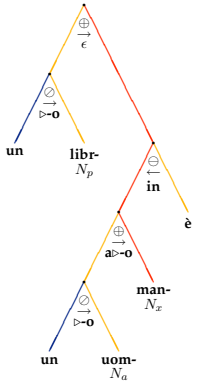


Figure 2.53: The adtree of *Un libro è in mano a un uomo* (48).

Figure 2.54: The Agent is trajector while the Patient is a blue landmark (49).

- (49-en.) A man is ⊗(with) a man.
- (49-it.) Un uomo è ⊗(con) un libro.
- (49-tu.) O kitap⊗-lı adam-dır.

Again, in example 49 the English and Italian adtrees are very similar (see Figure 2.55. Adpositions *with : con* are false, as the bivalent verb is *to be*

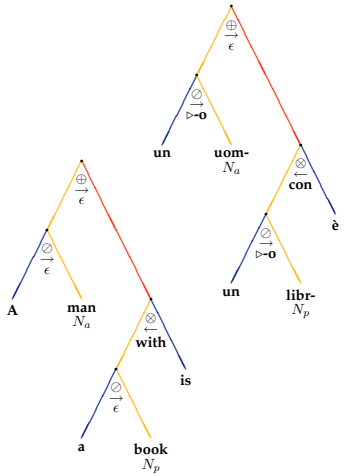


Figure 2.55: The English and Italian adtrees of *A man is with a man* (49).

with : essere con, a situation already explained in section 2.9.1. In contrast, the Turkish adtree is far more simple (see Figure 2.56). Example 49 is built morphologically as follows:

O	kitap-	-lı	adam-	-dır
That	*book*	*his*	*man*	*is*

Interestingly, the Turkish language use the determiner *O* as the syntactic subject *S*, instead of the Agent *adam-*, which is part of the second valence, unlike English and Italian.

2.11.4 The relation of possession

Let us see how the convention for PhAdS acts in NLs like Italian and English, while in Turkish is not so heavily needed. In conceptual space 50, the

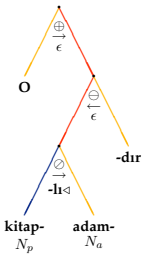


Figure 2.56: The adtree of *O kitaplı adamdır* (49).

book B is the tr while the man A is its owner, i.e., the Agent A is in a non-dimensional retroapplicative relation with B (in fact, it is the owner even *in absentia*). In this case, the three NLs use different strategies:

- (50-en.) The man∅('s) book.
- (50-it.) Il libro ∅(di) quell'uom-o.
- (50-tu.) Adam∅(-ın) kitab-ı.

Example 50 is shown in Figure 2.57. The strategy of English is shown in

Figure 2.57: The Patient as tr with a blue retroapplicative lm.

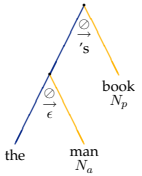


Figure 2.58: The adtree of *The man's book* (50).

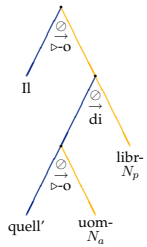


Figure 2.59: The adtree of *Il libro di quell'uomo* (50).

Figure 2.58: two determiners – i.e., Slash adpositions – are encapsulated one into the other symmetrically. In contrast, Italian should put three Slash adpositions into the syntagmatic axis to solve the deixis (see Figure 2.59, while Turkish need only two, as shown in Figure 2.60

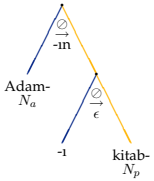


Figure 2.60: The adtree of *Adamın kitabı* (50).

2.11.5 The diathesis transformation active-passive

The transformation from active to passive voice is very important. In example 51 the Agent is tr, while the Patient is lm, and the action is expressed actively – see Figure 2.61. There are no circumstantials, so there is no need

Figure 2.61: A standard *VO* sentence, the verbal form in active voice (51).

of bold in adtree drawing. Vice versa, in example 52 the same content will expressed passively (Figure 2.62). Here, the Agent will be expressed by a

Figure 2.62: A standard VO sentence, the verbal form in passive voice (52).

circumstantial, therefore its nodes will be printend in plain. The three NLs express example 35 as follows:

- (51-en.) The man read-s $\ominus$ the book.
- (51-it.) L'uomo legg-e $\ominus$ il libr-o.
- (51-tu.) Adam kitab$\ominus_{\text{I}}$ oku-yor.

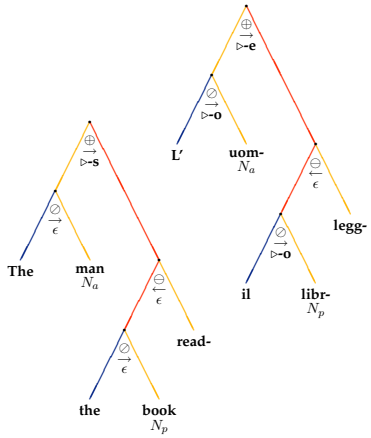


Figure 2.63: The English and Italian adtrees of *The man reads the book* (51).

Figure 2.63 shows how English put concept in the conceptual space: the verbal form *reads* is rightmost, as it is known, while *the man* (Agent) is leftmost, as it is the new information. Finally, *the book* is below *the man* being the second valence. Italian conducts analogously, except for the explicit marker in the Plus relation. Vice versa, the Turkish adtree is simpler, as it needs no determiners (i.e., there are no Slash relations; see Figure 2.64).

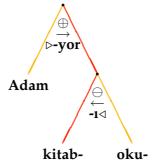


Figure 2.64: The adtree of *Adam kitabı okuyor* (51).

Example 52 shows the same phrase in the passive diathesis (Figure 2.62). The first step is to define the valence of the verb at passive: in all three NLs the lexemes became monovalent (this information is available by dictionary). Therefore, tertiary lms (i.e., circumstantials) should be provided if we want to express explicitely Agent A.

- (52-en.) The book is read ⊘(by) the man.
- (52-it.) Il libr-o è lett-o ⊘(da)-ll'uom-o.
- (52-tu.) Adam⊘(-ın) oku-duğ-u ⊘ bir kitap.

Example 36 is similar to example 23 (see the previous section). In English, the N_p is moved as the S, at the immediate left of V, while N_a became a circumstantial with a Slash adposition, which is opposite to Plus. Now, the Plus relation is head by the auxiliary verb *is* (see Figure 2.65). The adtree of Italian is identical to the English one, except for the leaves. The case of Turkish is different (see Figure 2.66). Example 52 in Turkish is built morphologically as follows:

adam-	-ın	oku-	-duğ-	-u	bir	kitap
man	*of*	*read*	+REL,+PASS,+PRES	*his*	*a*	*book*

The relative phrase *Adamın okuduğu* (literally, 'man write') is a circumstantial of the verbless phrase *bir kitap* (literally, 'one book'; compare with the Chinese Mandarin *hěn guì* seen in example 35, Figure 2.31). The complex morpheme *-duğ-* introduces a relative clause and at the same time modifies the verbal node *oku-*.

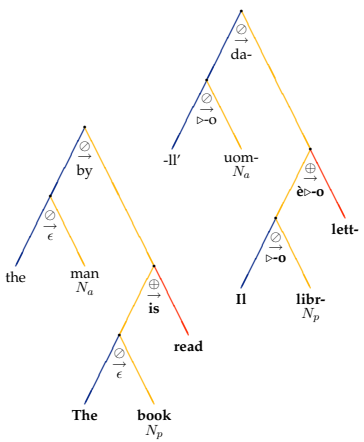


Figure 2.65: The English and Italian adtrees of *The book is read by...* (52).

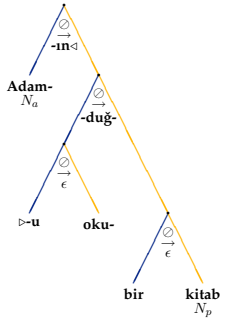


Figure 2.66: The adtree of *Adamın okuduğu bir kitap* (52).

Notes

[1] I am partly following section 2.2 of Schubert [1987] for the dichotomy constituency/dependency.

[2] See Chapter 3 of Schubert [1987] for the reception of Tesnière's work.

[3] However, an interesting proposal for an efficient dependendy-based parser is Topological Dependency Grammar (TDG) proposed by Debusmann [2001], which addresses exactly this kind of problems.

[4] This intuition is taken from Silvio Ceccato's works, in particular Ceccato [1961a] and Ceccato [1961b]. The main difference between adgrams and Ceccato's approach is the typology of possible relations. In fact, Ceccato used a semantic network where arcs are tagged with the appropriate relation. Consequently, the numer of possible tags increased enormously. In contrast, adgrams have only four types of relations, as I will present in the following sections.

[5] The English preposition *without* is clearly formed by *with* and *out*; nonetheless, the degree of grammaticalisation is so high that a morphological analysis is useless (see section 2.4 for details about nominalisation and grammaticalisation phenomena).

[6] The examples in Figure 2.1 are taken from the Italian translation Brøndal [1967, 197–204], as the original Brøndal [1940] is very difficult to find.

[7] Mnemonically, readers can easily remember the abbreviation 'tr' – the first two letters of both 'trajector' and 'triangle'.

[8] Under a phonological perspective, the tr often is revealed by a more relevant pitch than the lm. This line of reasoning is of maximum interest; however, it is out of the scope of the present dissertation.

[9] According to Langacker, the dichotomy trajector/landmark is conceptually distinct from the figure/ground distinction, as the first pertains the focus of attention, while the second one pertains the foreground-background perspective (see Croft and Cruse [2004, :58, note 5]). However, this nuance is not important in adgrams.

[10] For examples, cognitive linguists put themselves questions like: why in Italian we *andiamo dal dottore* (lit. 'we go from the doctor') when in English *we go to the doctor* (**andiamo al dottore* in Italian)? These questions can be interesting for the encyclopædic scope of (lexical) semantics, but they fail completely to build a coherent system. This kind of differences belong to the non-compositional part of the semantic structure, i.e., they belong to content, not to the conceptual space (see Introduction). As argued by Ferdinand de Saussure, each NL is an autonomous system in respect of its phonological and morphosyntactic relations. Therefore, it makes no sense to analyse prepositions cross-linguistically prima facie, i.e., without having established previously a robust linguistic system.

[11] The example scenario is made by sentences and pictures kindly given by Pennacchietti.

[12] See, for instance, Pennacchietti [2006b] for an alternative explanation in Italian of the same line of reasoning.

[13] For the purposes of this dissertation, adgrams are made on nominative-accusative NLs, i.e., ergative-absolutive NLs like Basque are out of the framework.

[14] See Chapter 3 and 4 in Taylor [1989] for a detailed account of the (dis)advantages of the classical and the prototypical approaches to linguistic categorisation.

[15] Source: Ramonet, Ignacio. 'Africa says no – and means it'. *Le Monde Diplomatique*. English edition. 2008, January. I will use texts from this international review because it is published on the web in 17 languages, among others, Esperanto. See chapter 4 for details.

[16] When referring to single NL grammars, I will use terms from the grammarian tradition when they are well-established. In my view, traditional grammar is often clearer in defining morphosyntactic phenomena than new proposals. Of course, these terms cannot be generalised, i.e., applied cross-linguistically, without taking great care.

[17] I should thank Pennacchietti for giving me this (unpublished) classification of English.

[18] When it is used prepositionally, *she rolled off the bed*, is neither as in *the wedding is off* nor *he ran off*, where it is used as a circumstantial, see section 2.9 hereafter for details.

[19] This treatment of the English preposition *with* is an oversimplification. In fact, in a sentence like *The farmer shot the rabbit with a hammer* the role of *with* is far different (more precisely, it is a Slash ∅). I will discuss this case speaking about the adpositional space of Esperanto in the second part, which distinguish clearly between the two roles of *with*; in NLs like English or Italian these cases can be solved only by collocation.

[20] Ceccato [1961b] talked about *correlazioni*, 'correlations', for what I call adpositions, and *primo correlato* 'first correlate' and *secondo correlato* 'second correlate' the two groups involved by the adposition. Ceccato's program of research has been presented in Italian in Ceccato [1961a]. His major English contribution is Ceccato [1961c] in Ceccato [1961d]. After the famous Report by ALPAC [1966], this research program ceased. For an evaluation of Ceccato's Correlational Grammar, see Glaserfeld [2001] in Hutchins [2001].

[21] More precisely, example 15 entails example 14, because 14 is true whenever 15 is true. Consequently, 14 and 15 are synonymous, i.e., even if their conceptual spaces differ, they evoke the same content. See Chapter 3 for the notion of entailment and synonymity.

[22] In reality, the angle is a bit less than 45 degrees, for typographic needs which will be obvious when adtrees will become very complex.

[23] It is important not to identify groups with the chomskian Noun Phrase (NP) or Prepositional Phrase (PP). In fact, any minimal adtree has an adposition as head – it can be a zero adposition if it is signed by collocation only; nonetheless, it *is* an adposition. Hence, any minimal adtree should be called a Prepositional Phrase in Chomskyan terms, and this is clearly useless.

[24] As said in the opening of this chapter, I will retain the tesnerian notion of valence; readers should care that cognitive linguists use the term 'valence' in a different sense (see the glossary in Taylor [2002]). Note also that Anna Wierzbicka is a notable exception: she distinguish dictionary and encyclopædia. See at least Wierzbicka [1988].

[25] The analysis of actants is important to perform what Tesnière called 'metataxis', i.e., the set of rules necessary to transform adtrees, for example for translation purposes. I will come back to this topic in the third part.

[26] I will partly follow the taxonomy offered by Taylor [2002], section 21.3, partly the used by Pennacchietti, who adheres to the original model proposed in the classic contribution by Fillmore [1968] in Bach and Harms [1968]. In particular, I consider Agenthood as a category both with animate and inanimate entities, because they are functionally equivalent. An immediate corollary is that an Instrument can be animated as well as inanimated. Finally, I do not consider the ZERO category proposed by Taylor [2002] really useful for my purposes.

[27] Porphyry was a third-century scholar who wrote an exegesis of Aristotle's notion of genus and species, *Isagoge*. In this work he introduced a taxonomy with the aim of classify any entity, from God to nature. Its taxonomy is structured as a binary tree which deeply influenced Western philosophy. If the reader wants to delve into the topic, I suggest starting from Eco [1993]. Note that Tesnerian trees are not Porphyrian at all: they resemble the trees used by the German philologist Karl Lachmann, mostly used in stemmatics [Timpanaro, 1990].

[28] In Italian, there are three different paradigms of 'weak verbs', signed by the so-called 'thematic vowel' in the infinitive: *-are*, *-ere* and *-ire*. The signatures of the three different paradigms of weak verbs are equivalent as the three paradigms are symmetric.

[29] I borrowed the following examples by Langacker, but I grouped them differently.

[30] In the case of circumstantial actants, it is the adposition which give the right character. In examples 22-26, *at* is a punctual locative adposition, so the corrispondent semantic role will be Place, following the taxonomy by Taylor [2002]. However, for the purposes of this dissertation, actants play a role mainly to clarify how valence works, so I will not delve into the complex and debated topic of the 'right' or 'universal' taxonomy of semantic roles. In my view, these should be applied only when they are needed to disambiguate nuances beyond morphosyntax, i.e., I use Occam's razor about the taxonomy of semantic roles.

[31] In Pennacchietti's terminology, primary and secondary lms are collectively defined as 'inferior circumstantials', while circumstantials are called 'superior circumstantials'. These

terms are appropriate if the reader looks at the structure of adtrees. However, I prefer to speak about primary, secondary and tertiary lms.

[32] Here, I'm following Chapter 106 of Tesnière [1959] and the symmetrical Chapter 80 of the partial Italian translation Proverbio and Trocini Cerrina [2001]. As far as I know, unfortunately by now there is not a translation into English of this classic work.

[33] Conversely, note that it is possible to have sentences with two datives to express the second and the third actant, as in the Latin sentences *id est mihi* {*gaudio*|*cordi*} ('I'm happy for this|I take it to heart').

[34] I borrow the following examples from Tesnière [1959] and the Italian translation from Proverbio and Trocini Cerrina [2001].

[35] If some native speaker or language expert of Italian or French can find a relevant example, please drop me an email.

[36] This notation is a slightly adapted version of the one presented in Pennacchietti [1981a, 304].

[37] I acknowledge the participants to a seminar of mine about the basics of adgrams for their provoking questions. It was given in the academic year 2007-8 at the University of Insubria, Varese.

[38] I prefer the whorfian term 'stative' instead of 'nominal' or 'noun', in order to avoid possible confusion with the well-established Chomskyan term Noun Phrase (NP), that I already discuss in the opening of chapter 2.

[39] This example is extracted from Taylor [2002, 381].

[40] I call the phenomenon 'mirroring' and not 'reflexion' because I have already used that terms to clarify how adspaces are structured; however, it is exactly the same mechanism.

[41] I will deal with Esperanto relative clauses in the second part of this dissertation. The topic is so complex that a full dissertation should be devoted for a cross-lingual description of relative clauses in terms of adgrams. For a preliminary treatment, see Pennacchietti [2007] in Venier [2007].

[42] Symmetry in adgrams has an apart status: I decide not to represent *cats and dogs* adtree with a shortcut, for the reader's sake. I will come back later on this very important topic.

[43] The prototypical correspondent phrasal adpositions in Italian are *ma* and *però*. There are others, but less frequently used, such as *peraltro*, *eppure*, *senonché*. Note that they are all results of grammaticalisation, as seen in the Introduction. This fact was brillantly shown by NLP (Neuro-Linguistic Programming), in particolar in the classic work by Bandler and Grinder [1975].

[44] I will explain the Indo-European correlatives like the English series *wh-* or the Latin/Italian series *qu-* in the second part of the dissertation, through the example of Esperanto.

[45] It can also be said that *and* and *that* are two signatures of different adpositions; that is equivalent.

[46] I extended the example in Tesnière [1959, 337] and Proverbio and Trocini Cerrina [2001, 195] with the English version.

[47] Contra Tesnière, these two actants are not equal: in NLs like English or Italian the actant on the left is more salient than the actant on the right.

[48] This example is a slightly different version of the one provided by Proverbio and Trocini Cerrina [2001, 191], where it explains zero junctives.

[49] I wish to thank the lecturers in Chinese and Japanese of the Dipartimento di Orientalistica, University of Turin, Italy, for these examples.

[50] I will explain in the second part that surprisingly there is a specialised morpheme in Esperanto which the same function: *ki-*. See section 5.3.9 for details.

[51] Also Korean shows two morphemes alike. It is not a goal of this dissertation to build the adgram of Japanese. For Japanese, I based myself on Yamasaki [2000], who presents Japanese to Esperanto speakers.

[52] I am in debt with Pennacchietti who gave me the example – especially the Turkish version – which he explained to me, and all the pictures.

[53] I wish to thank Pennacchietti for giving me the adspace of Turkish.

[54] Again, if some native speaker or language expert of English or Turkish can find a relevant example, please drop me an email.

Chapter 3

Dictionary and lexicon

The lexicon is the set of meaningful units of a given NL. In cognitive grammar terms, while adpositions give the co-text grounding, the meaning belongs to lexemes. In other words, even if every morpheme is a brick of the language building, only lexemes are specialised in conveying meaning. Chapter 2 showed adtrees, the scaffolds of meaning, while in this chapter the building will be completed, presenting how the lexicon is made and how it is put into adtrees.

Adgrams fully endorse the conceptualist approach already presented in the introduction: i.e., contra Dewey, Wittgenstein and Quine, there *is* more to the meaning of an expression than the overt use that we make of the expression.[1] I argue that language – even considered only as a skill – cannot be reduced to a mere observation of behaviours: human beings are more intelligent than Weizenbaum's Eliza. It is true that there are semantic and pragmatic aspects that can be solved only in terms of encyclopædic knowledge, and I will fully acknowledge the limits of adgrams in that sense. The error made by the philosophical line followed by Quine is that encyclopædia can explain *every* linguistic knowledge, i.e., language (or, at least, meaning) is strictly reduced to use – we could call it a 'panencyclopædic' approach to language. Within the cognitive grammar approach, only Wierzbicka [1988] insists to distinguish **encyclopædia** versus **dictionary**, which have different properties – see Table 3.1. This distinction is still valid in adgrams. As said in the introduction, semantics cannot be captured by a semantic metalanguage that evaluates meaning in terms of truth and falsity: adgrams are agnostic with respect to world models. Encyclopædia does not need to be compositional. Meaning is made by the encyclopædia *and* the dictionary, but adgrams deal only with the latter – except of valence and actants. Therefore, each dictionary is language-dependent: in this chapter I will give examples from different NLs.

Lexical semanticists often adhere either to the monosemic analysis, starting from De Saussure [1970], or to the polysemic analysis [Katz and Fodor,

Table 3.1: Encyclopædia and dictionary in comparison

	encyclopædia	**dictionary**
Scope	grounding	coherence
Dependency	world model	language system
Structure	mostly non-compositional	mostly compositional
Elements	valence, actants	morphemes, collocation

1963]. For example, in the polysemic approach, the different meanings of the lexical entry *bachelor*, such as 'an unmarried adult man', 'a person who has taken a first university degree', and 'knight serving under the banner of another knight', are viewed as disjoint, while in the monosemic method a more abstract meaning, 'unfulfilled in typical male role' is found, being the prototypical meaning – see Kornai [2008, 488-489] for details. Following the principle of linguistic adherence presented in the introduction, adgrams are inclined to the monosemic approach, at least within a single NL. In fact, it is perfectly coherent and obvious to separate the different meanings of a single lexical entry like *bachelor* when performing a translation. For instance, in Italian the prototypical use is rendered as *celibe* while the holder of a BA degree is usually translated to *diplomato* – see chapter 8 for more details.

But how is this inventory of meaning structured? The dictionary is made by the lexicon and the adspace. While the adspace – the set of adpositions of a given NLs – does participate into the construal process of meaning building, the lexicon as a whole is vehicular of semantics. The lexicon is made by the infinite set of lexemes and the finite set of affixes. Traditionally, morphemes are divided into **free morphemes**, like *farm*, *rabbit*, *roof*, *slope*, *gent*, *down*, *up*, which can stand alone in the linguistic chain, and **bound morphemes**, which should be always attached to free ones, like *-er*, *-s*, *-ly*, *-wards*.

I have just talked about 'lexical entries', which is synonym of 'lexeme' where I refer to the structure of the dictionary. For the moment, I call **lexeme** a no further analysable free morpheme. For instance, *farm* is a lexeme while *farmer* is a word made by two morphemes: the free morpheme *farm* and the affix *-er* (more precisely, the suffix *-er*). In adgrams a **word** is simply a given combination of two or more morphemes.

The dichotomy free versus bound is very useful and functional for morphology, so I will retain it, but it is not sufficient for depicting a good inventory. Linguists often consider the core of the lexicon made by free morphemes, while the various grammar relations should belong only to bound morphemes. But there are a lot of morphemes, like *the*, *was*, *by*, which are free and clearly have grammar functions (the morphemes used as exam-

ples until here are limited to example 1, presented in the introduction). In other words, this dichotomy is good for the morphological mechanics but it is of no use for depicting semantics.

In adgrams, the key concept for lexical semantics is the notion of **grammar character**, which will be fully analysed through other key concepts in the next sections. More specifically, from Whorf I borrow the notions of **modulus** and **signature**. Afterwards, these two concepts will lead from one side to **transference**, a concept introduced by Tèsniere, and to the other side to **grouping**, which is borrowed from the Chomskyan tradition.

3.1 Grammar character, moduli and signatures

We need general, cross-linguistic criteria to find the inner structure of morphemes. Looking to the linguistic research into universals of language, there are two 'conceptual archetypes' (Langacker) that can be found in any NL: things and events. Of course, each NL shapes in its own way things and events, that traditionally are signed with the grammatical categories of 'noun' and 'verb', which are two opposed poles. The main problem of traditional grammatical categories like noun, verb, adjective, adverb, passive voice, etc., is that they are usually considered 'fixed into the lexeme', because they derived from classical tradition, i.e., from ancient Greek and Latin grammarians.

In what follows, I will argue to solve the problem of grammatical categories, partially following a proposal by Whorf, almost completely neglected by theoretical linguistics, as far as I know.[2] Whorf refers to the familiar 'parts of speech' of most European NLs as **selective grammar characters**, i.e., their membership is fixed contrasted with the opposed polarity, without any regard to collocation. For example, in English, there are some words that are selective **verbs**, such as *reduce*, *survive*, *perplex*, *magnify*, *reciprocate*, while other words are selective **nouns**, as *instrument*, *elephant*.[3] I call 'noun' and 'verb' only lexemes whose grammar character is clear by selection. In fact, in such cases, an English speaker does not need to collocate them into the co-text, i.e., on the syntagmatic axis, in order to know their role in the phrase: he or she knows it by selection.

However, most English words cannot be marked selectively as verbs or nouns without co-text. Some examples are: *head*, *hand*, *stand*, *walk*, *exchange*, *sight*, *skin*, *weave*, *dog*, *surrender*, *massage*. In these cases, we should to talk about **modulus of stativation**, when the morpheme acts as a noun, and **modulus of verbification**, when it acts as a verb. For instance, I can say that *walk* can be stative in *let's take a walk* and verbified in *I walked across the street*.[4] Therefore, it is better to talk about 'grammar character' instead of 'grammar categories', as the term 'character' refers to a property, which implies that it can be present or absent, while 'category' refers to something

Table 3.2: Two stativation and verbification signatures in Hebrew

	e-e		*ā-a*	
b′r′k	*berek*	'knee'	*bārak*	'he kneeled'
d′r′k	*derek*	'road'	*dārak*	'he marched'
g′b′r	*geber*	'strong man'	*gābar*	'he was strong'

fixed, i.e., a morpheme *always* belongs to that category. Stativation and verbification are fundamental moduli and, as we will see hereafter, they indicate the lexical heads of their respective groups.

3.1.1 Moduli and signatures

In some NLs, there are clear morphological patterns that indicate the presence of moduli. These are called **signatures** for the moduli. Signatures form paradigms. Let me use see some examples from Whorf [1945]. In Hebrew, there are two signatures for stativation and verbification moduli, consisting of given vowel-consonant sequence patterns: *e-e* for stativation and *ā-a* for verbification (see Table 3.2). In most NLs, there can be more signatures for the same modulus. For instance, for the modulus [±PLURAL] in English we have more than one signatures, where the most productive ones are *-s* and *-es*, as in the couples *rabbit : rabbits*, *fish : fishes*. I call *-s* and *-es* **overt signatures** of the English plural. Signatures are not always overt: they can also be covert. For instance, in the case of selective nouns and verbs, signatures are carved in the very lexeme. Selection is the simplest case of covertion; however, sometimes covertion is indicated in more complex way. Let me explain through an example:

- (53a.) A fish appeared.
- (53b.) ϵ-Fish appeared.
- (53c.) The fish will be plentiful.

In example 53a the presence of the article *a* denotes that a single, particular fish was seen, while in 53b the *absence* of the article (marked by the sign: ϵ) denotes *plural*. In 53c, the signature *the* of the [−PLURAL] modulus is overtaken by the selective adjective *plentiful*. If in 53a the signature of the [−PLURAL] modulus is overt, in 53b and 53c the signature is covert. There is always a **covert signature** when there is no visible signature, i.e., no morpheme specialised for that syntactical function. Alternatively, it is possible to say that plurality is signed by a zero morpheme.[5]

3.1.2 Equivalence, synonymy and homonymy

Equivalence and synonymity refer to the paradigmatic axis. Sometimes there are two different signatures with the same collocation, i.e., they can be substituted without affecting meaning. In English, the series *don't*, *doesn't*, *can't*, can be substituted with *do not*, *does not*, *can not*.[6] For these cases, I will say that these signatures are **equivalent**. When signatures or lexemes are equivalent, I call them **synonyms**: in adgrams, synonymy is defined only in terms of language-internal relations within a single NL, without invoking semantic notions like entailment (see Taylor [2002, 190] for this last approach).

Perfect synonymy is very rare indeed: mostly, there is a specialisation, i.e., one signature can be collocated in some cases while the other cannot, as for instance in the signatures of the English plural *-s* and *-es*. In Italian, I have found only two pairs of perfect synonyms: *tra : fra* (two signatures of the same preposition, approximately 'between') and *termosifone : calorifero* ('radiator': the first word is made through Greek morphemes, the second one through Latin ones, and they are exactly symmetrical). What it is important is that their *collocation* should be the same, i.e., they can be substituted in every instance of the syntagmatic axis. In case of imperfect synonymy, the substitution can happen only in certain cases: a collocation analysis should be performed, in order to find the moduli involved that permit and block the substitution. It does not matter if the two signatures do not affect meaning: as in the example of English plurals, the reason behind the specialisation can be phonological. In other words, the moduli involved in blocking the substitution are phonological in nature. Adgrams do not take into account the *causes* of these blocks: the dictionary will register it when describing transfers (see below for details).

Finally, as we will see in chapter 8, collocation will be used not only for synonymy but also for cross-linguistic equivalence, i.e., for (machine) translation.

As we have seen before, the English morpheme *-s* is a signature of the modulus [+PLURAL] and it inherits stativation: i.e., *-s* can only be attached immediately after a stative lexeme. However, not every English morpheme *-s* is a plural marker, as shown in Table 3.3, based on the recent dictionary of English affixes by Stein [2007]. the English final *-s* can sign very different moduli. These five morphemes *-s* are signature of five different moduli: they are **homonyms**. Homonymy is the complementary phenomenon of equivalence. Normally, in order to achieve a good mastering in the target NL, a lot of cognitive effort is spent in solving homonyms, through collocation analysis.

In Table 3.3, the signs I>O and O>E indicate transfers (see below for a detailed explanation of transference). For the moment, let us concentrate on the last constrastive couple: *Monday : Mondays*. What I want is to show

Table 3.3: English homophones of *-s*

signature		*modification*	*constrastive couple*		
-s	:	[+PLURAL]	*hat*	:	*hats*
-s	:	[+PET,+NICK]	*dog*	:	*doggies*
-s	:	[+3rd PERSON]	*(to) beg*	:	*(he) begs*
-s	:	I>O	*(to) gape*	:	*(the) gapes*
-s	:	O>E	*Monday*	:	*Mondays*

how synonymity can exist not only between a single pair of lexemes, but also between groups of lexemes of the same type (again, see below for a detailed description of grouping). I will do it through example 54.

- (54a.) The restaurant is opened $\ominus$(Mondays).
- (54b.) The restaurant is opened $\ominus$(on Mondays).
- (54c.) The restaurant is opened $\ominus$(each Monday).
- (54d.) The restaurant is opened $\ominus$(Monday evening).
- (54e.) $\ominus$(The last Monday of each month) the restaurant is opened.

The same construal can be built through a lot of different linguistic strategies see circumstantial lexemes in brackets): *Mondays*, *on Monday*, *each Monday* in examples 54a-c are equivalent in the sense given above, while example 54d and 54e present the construal with more details.

If there is no clear signature, i.e., no morphemes attached to *Monday*, we should turn our attention to collocation to see which is the grammar character of the lexeme *Monday*. In fact, in example 54d the collocation of *Monday* next to *evening* and immediately after the verb *opened* allow us to conclude that it is circumstantial, while in example 55e *Monday* is preceded by the article *the* and so it retains its original stative modulus. Example 54 showed that morphology and syntax are two sides of the same coin. NLs have complex ways of performing these modifications, and this is why years of learning are needed to acquire an acceptable level of performance in any NL.

3.1.3 The fundamental moduli

Another important property of moduli is **inheritance**: some moduli can be activated only within a hierarchical superior modulus. For instance, in English, the modulus [±PLURAL] can be actived only if the modulus STATIVATION is active.[7] In contrast, in Italian the modulus [±PLURAL] must be

Table 3.4: The fundamental moduli

STATIVATION	→	O
ADJUNCTIVATION	→	A
CIRCUMSTANTIATION	→	E
VERBIFICATION	→	I

actived within both the modulus STATIVATION and VERBIFICATION. When a modulus is spread out over the boundaries of the fundamental moduli we have **agreement**, which is the phenomenon complementary to inheritance. For example, in the Italian sentence *quella bella ragazza ballava bene* ('that nice girl danced well') the final morpheme *-a* in *quell-a* and *bell-a* is a signature of gender and number agreement between the selective adjectives and the selective noun *ragazz-a*, which also ends in *-a*.

Which are the hierarchically highest moduli? As we have seen previously, the polarity between STATIVATION and VERBIFICATION is a language universal. Similarly, there are always two symmetrical moduli whose function is to modify, specify, or concretise, stativation and verbification. I call **adjunctivation** the modulus which modifies things.[8] Paraphrasing Tesnière [1959], who did not talk explicitly of moduli, I will call the modulus which modifies the events **circumstantiation**. Lexemes where the adjunctivation modulus is applied either by selection or by collocation will be called **adjuncts**. Adjuncts include adjectives like *yellow*, *big*, and determiners like *a*, *the*, *every*. In contrast, if the circumstantiation modulus is selective, that lexeme will be called **adverb**; vice versa, if the circumstantiation modulus is applied by collocation, that lexeme will be called a **circumstance**. [9] This choice is due to the fact that "true" adverbs are very rare, at least in the NLs I know, and so I prefer to dedicate this term to a restricted set of particular lexemes.

Of course, each NL has its own strategies to mark these moduli by one or more (c)overt signatures: they are language universals because at least one strategy per modulus exists in every NL. Hereafter, I will refer to these four as **fundamental moduli**. Typographically, fundamental are signed in small capitals without square brackets, unlike the other moduli. All other moduli are hyponyms of the fundamental ones and they are language-dependent. The fundamental moduli give the **grammar character** to each lexeme and to groups of them: the grammar character is the cognitive representation of a given set of lexemes put onto the syntagmatic axis in order to construe the scene of meaning in the mind.

Since fundamental moduli will be extensively used, I will use the shortcut originally by invented Lucien Tesnière in his French work *Elements of structural syntax*, as in Table 3.4, already in Gobbo [1998, 60]. Readers can

Table 3.5: Tesnière levels of analysis and adgrams in comparison

Tesnerian proposal	**traditional terms**	**terms in adgrams**
translation du premiere degré	prépositions	basic-level adpositions
jonction	conjonctions de coordination	phrasal adpositions
translation du second degré	conjonctions de subordination	phrasal adpositions

think metaphorically at O and I as two poles, and at A and E as their respective magnetic field.

3.2 Transference

It is worth remembering how Tesnière [1959, 387] distinguish the three levels of analysis of structural syntax, and how they are interpreted in adgrams (see Table 3.5). The basics of the Tesnerian approach is dependency. Tesnière put "connexion" (literally 'connection') at a word level, i.e., connection is the morphosyntactical relation between words, the mechanism of determination of hierarchies being dependency. I have shown in Chapter 2 how adspaces solved the syntactic problem of dependency through the dichotomy trajector/landmark, taken from cognitive linguistics, and that adpositions can be applied at two levels: basic-level and phrasal – see section 2.10.

What I want to introduce at this point is a further aspect of the Tesnerian "connexion", called in French "translation" – in English 'transference'.[10] **Transference** is a very important phenomenon that explains the inner structure of the lexicon. The phenomenon of transference refers to grammar character change: it is the interface between adspaces and the lexicon. Put it differently, transference answers the question: how do lexemes change their fundamental modulus? It is impossible to give a complete answer to this question, because a full list all moduli of every single would be needed. Here, I limit myself to explain the methodology, while in the second part I will give a broader instance, which will be implemented in the third part. In the following example I give the table of correspondance of NL couples, which also is an example for (machine) translation purposes – see chapter 8.

The first step is to identify the fundamental modulus of the lexeme. This identification can be done through the collocation analysis, with a special regard to the bound morphemes at which it is attached. In fact, bound

Table 3.6: How transference works for A-lexemes in different NLs

transf.	*English*	*Italian*	*French*	*German*	*Turkish*
A	long	lung-o	long	lang	uzun
A>O	length	lungh-ezz-a	longu-er	Länge	uzun-luk
A>E	long	lung-amente	longu-ement	ent-lang	uzun
A>I	length-en	al-lung-are	(r)al-long-er	ver-läng-ern	uzatma-k

morphemes, as well as free morphemes crystallised in idiomatic expressions, often have only the function of grammar change markers. In case of doubt, a cross-linguistical analysis would help the identification.

The fundamental modulus of the English lexeme *long* is adjunctivation. An equivalent expression to say that is: '*long* is an A-lexeme', which is more concise. This identification is not obvious, as in *Liza has long red hair* the lexeme *long* is an adjunct, while in *it rained all day long* the same lexeme is a circumstance. A hint can be found by the fact that *long* as a circumstance is marked by the modulus [+TIME], while – as an adjunct – it can be marked by the modulus [+SPACE], as in the previous example, or hence by the modulus [+TIME], as for instance in *a long career*. Table 3.6 shows how transference works in different NLs in comparison. Italian and French are Romance languages, English and German are Germanic (at least morphologically), while Turkish is not Indo-European: this should offer a broad example in terms of typological variety. The German and English strategies of the transference A>O is ablaut, while English transfers A>I with a mixed strategy of ablaut and derivational suffix. Italian are quite similar, using suffixes for the transfers A>O and A>E, while for A>I they use circumfixes, like German. Note that, unlike the other NLs under exam, the German *lang* is marked as [+SPACE,+FAR],[−TIME], while *lang-wierig* is reserved for the complementary [−SPACE],[+TIME] (as a side note, [+FAR] is activated only within its hierarchical superior modulus [+SPACE], and this is the reason why they have been grouped within the same square brackets). Turkish is different, as *uzatma-k* derives from *uzatma*, which is O in character, meaning *extension cable*. This means that two different paradigms are involved: each NL has its own semantic map, which is not necessarily burdened in the same way; the more strcucturally and typologically distant, the more semantic maps are distant indeed. Nonetheless, interestingly, Turkish has a suffix for the transference A>O, *-luk*, like the Italian *-ezza*.

Table 3.7: An example of suppletive transference

transf.	*English*	*Italian*	*French*
A	calm	calm-o	calm-e
A>O	calm	calm-a	calm-e
A>E	calm-ly	*calm-amente con calma	calm-ement
A>I	calm (down)	calm-are	calm-er

Table 3.8: Specimen of the transference A>E in English and Italian

English			*Italian*		
sudden	:	sudden-ly	subitane-o	:	subitane-amente
high	:	high-ly	alt-o	:	alt-amente
tranquil	:	tranquil-ly	tranquill-o	:	tranquill-amente
sudden	:	sudden-ly	subitane-o	:	subitane-amente

3.2.1 Suppletion and idiomatic expressions

The most expensive strategy, in cognitive terms, is **suppletion**: whenever a NL has a block in one of its paradigms, it should fill the gap with a suppletive lexeme. In the linguistic literature, the concept of suppletion is usually kept into the verbal domain: the classic example is that the lexeme *went* is suppletive of the past form of the verb *to go*, i.e., it supplies **goed*. Given the role of transference, adgrams make an extensive use of suppletion. Let me explain through an example, where an A-lexeme is transferred in each other fundamental modulus, i.e., stativation, circumstantiation, verbification. Table 3.7 shows how subtle differences operate even in NLs which are siblings, like Italian and French. Note that Italian has a block in its paradigmatically transference strategy A>E 'add *-amente* to the A-lexeme' *only* with the lexeme *calm-*, in other cases highly productive – see a specimen in Table 3.8. Instead, Italian should use an idiomatic expression, *con calma*. Let me clarify this point using some lexemes presented above in an example.

- (55a-en.) Suddenly Liza walked calmly.
- (55b-en.) Suddenly Liza calmed her walk.
- (55a-it.) Subitaneamente Lisa camminava con calma.
- (55b-it.) Subitaneamente Lisa calmò la sua camminata.

This means that *con* (literally, 'with') in example 55 is a false adposition, because it has the same grammatical role of the English morpheme *-ly*. In fact, the expression *con calma* is completely crystallised, i.e., it cannot be changed, e.g., for gender and number: **con calmi*, **con calme* are ungrammatical. Alternatively, the Italian speaker can choose an imperfect synonym, like for instance *tranquill-amente*, with the problems of collocation already seen. **Idiomatic expressions** are special lexeme chains which have been crystallysed because of some need, in most cases a need of suppletion. In adgrams they are treated as a whole, even in their representation in adtrees.

3.3 Actants and adtypes

Figure 3.1 shows the adtrees of example 55 with the new information made explicit. By now, colours will be omitted being redundant: the signs for Plus, Minus, Slash, Times mark the adtree type (**adtype**), and the arrow remembers the direction of the relation trajector/landmark – the adverted reader can refer to Figure 2.8 in chapter 2. The results of the adpositional transfers – i.e., the grammar character changes performed by the adpositions – are signed below the lexemes. Now it is possible to define again adpositions with a lexical definition, by contrast.

> **adposition**: *a morpheme which does not have any grammar character per se but it gives the correct adtype (Plus, Minus, Slash or Times) to the current adtree.*

In many cases, derivation is the result of a long process of grammaticalisation (see again section 2.4 if needed). Now, it is possible to trace a grammaticalisation degree, based on grammar characters:

$$\text{I} \rightarrow \quad \text{E} \rightarrow \quad \text{O} \rightarrow \quad \text{A}$$

This means that the fundamental grammar character of a lexeme determines its level of grammaticalisation. However, transference can always change the grammar character of a given lexeme, with some precise limits. These limits give the correspondence between actants and adtypes, shown in Table 3.9. The table is ordered along the grammaticalisation of the grammar character, i.e., its characterisation. It is noteworthy that the original grammar character being transferred – indicated with x – is not relevant. The final characterisation can be the result of a complex chain of transfers, performed by derivational morphemes (see next section 3.4 for details).

3.3.1 Verbal adtypes

Lexemes with a verbal modulus active are always applicative: they can be either Plus, if the phrase is non-marked, or Minus, if the phrase is marked.

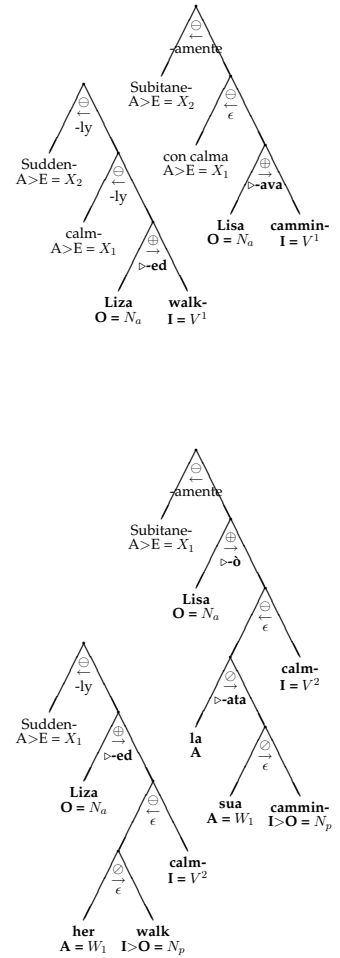


Figure 3.1: Some adtrees of *Liza*, *walk* and *calm* (55ab).

Table 3.9: The correspondence between actants and adtypes

final characterisation	*transference*		*actant*	→	*adtype*
verbal					
	y =I	=	$V^{\{1\|2\|3\}}$	→	⊕
	y =I	=	$V^{\{1\|2\|3\}}$	→	⊖
circumstantial					
	$x > y$ =E	=	$X_{1,\dots,n}$	→	⊖
stative (full)					
	$x > y$ =O	=	$\{N_a\|N_e\|N_p\|N_x\}$	→	⊕
	$x > y$ =O	=	$\{N_a\|N_e\|N_p\|N_x\}$	→	⊖
	$x > y$ =O	=	$\{N_a\|N_e\|N_p\|N_x\}$	→	⊗
	$x > y$ =O	=	$\{N_a\|N_e\|N_p\|N_x\}$	→	⊘
stative (void)					
	$x > y$ =O	=	$W_{1,\dots,n}$	→	⊘
	$x > y$ =O	=	$X_{1,\dots,n}$	→	⊖
	$x > y$ =O	=	$Y_{1,\dots,n}$	→	⊕
	$x > y$ =O	=	$Z_{1,\dots,n}$	→	⊗
adjuntive					
	$x > y$ =A	=	$X_{1,\dots,n}$	→	⊖
	$x > y$ =A	=	$W_{1,\dots,n}$	→	⊘

In this last case mirroring is applied – see again section 2.9.3 for details. Let me explain through an example in Italian:

- (56a.) Carlo è arrivato ('Carl has arrived').
- (56b.) È arrivato Carlo ('It is Carl, who arrived').

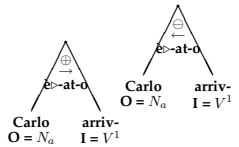


Figure 3.2: The adtrees of *Carlo è arrivato* / *È arrivato Carlo* (56ab).

Figure 3.2 shows the adtrees of example 56: in example 56a (left) the sentence is non-marked, while in example 56b (right) mirroring was applied.

3.3.2 Circumstantial adtypes

Circumstantials require only a shallow analysis, as they are clearly applicative related to the phrasal verb. For example, the presence of the English adposition *-ly* clearly states the circumstantial, even if the lexeme to which the adposition is nonsensical, like in the following example.

- (57.) He conducts mugglly.

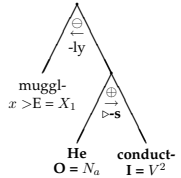


Figure 3.3: The adtree of *He conducts muggly* (57).

The lexeme *muggl-* does not exist in English; nonetheless, the syntactic relation is quite clear thanks to the suffix *-ly*.[11]

3.3.3 Stative adtypes

Actants are always attached to the trajector of their own adtree. As seen before in section section 2.7, actants are related to valence. Now it is possible to say that actants related to valences are either nominal per se (N stays

for 'nominal') or they have acquired a nominal, or better, a stative value. For example, pronouns like English *he* in example 57 have a stative value even if they are not selective nouns, but anaphoric **placeholders**. Pronouns like *he* and determiners like *this* can act as placeholder of actants. More precisely, they can be either **anaphoras**, if they stand for a previous stative lexeme in the co-text, or **cataphoras**, if they stand for a later stative lexeme. They are indicated with the sign t, which stands for 'trace'.[12]

It is noteworthy that adgrams do not deal with anaphora and cataphora resolutions, at least in this dissertation: a whole apart dissertation should be dedicated to this important topic. This is a limit of the current model of adgrams.

Some determiners also are placeholders. For example, *read* in example 58 and *this* in example 59b have both a stative value because of transference – in contrast, *this* in example 59a have not. Finally, note that *this* in example 59b acquires both stativation and an anaphoric or cataphoric value – i.e., it is a placeholder.

- (58.) Reading is enjoyable.

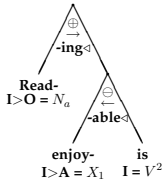


Figure 3.4: The adtree of *Reading is enjoyable* (58).

- (59a.) This book is mine.
- (59b.) This is mine.

Figure 3.5 shows the adtrees of example 59ab. A complete description of transference has been provided; note the actant values of the proper verb *to be*, i.e., the bivalent verb, which are syntactically symmetrical with the verb *to have*. In fact, examples 59a and 59c depicts the same construal.

- (59a.) N_p(This book) is N_a(mine).
- (59c.) N_a(I) have N_p(my book).

Stative actants can depend of adtrees belonging to *any* adtree type. Furthermore, if a stative lexeme does not bring an actant value, it will be filled with a **void actant**. For the most adverted reader, void actants have been

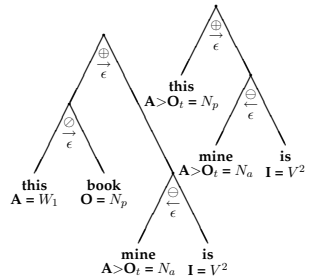


Figure 3.5: The adtrees of *This (book) is mine* (59ab).

already shown in section 2.8, without any explanation. However, void actants strictly depend on the adtree type; in other words, they have no pragmatic value, i.e., they are void (see Table 3.9, for the one-to-one correspondences).

In fact, in adgrams there is no use of actants for so-called 'thematic roles' like cause, goal, path, measure, etc. In my opinion, one of the most evident limit of such an approach is the lack of agreement of the inventory of these categories.[13] Moreover, it does not respect the principle of linguistic adherence already presented. In fact, these thematic roles are not actants, but phrasal arguments which are interpreted here as adtypes, depending on their adpositions. The four actants in use here – i.e., Agent, Experiencer, Patient, directly related to valence, plus Instrument – are very general and they are accepted by most linguists as cross-linguistically valid.

The void actants X and Y corresponds to the applicative (hot) relations, which normally are near the phrasal verb, i.e., the I-lexeme, while W and Z concern the retroapplicative (cold) relations. The numbers indicate the order of that adtype within the current adtree branch. They are used for building the construal and hence the adtree. For instance, in the English example 54a the order is as follows: *Liza walked* → *Liza walked calmly* → *Suddenly Liza walked calmly*. In fact, the sentence *Suddenly Liza walked* evokes a completely different content: in this last case, perhaps Liza was sitting, or maintaning an upright position without moving, etc..

3.3.4 Adjunctive adtypes

Adjuncts can be use as attributes or predicative: the first use is retroapplicative (Slash) while the second one is applicative (Minus). The English strategy of distinguish the two uses is collocation: the attributive use is put

before the stative lexeme while the predicative one is put thereafter.

- (60a.) That water is hot.
- (60b.) That is hot water.

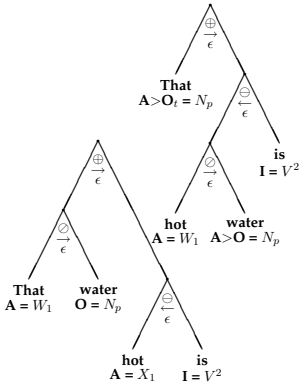


Figure 3.6: Predicative (left) and attributive (right) uses of adjuncts (60ab).

Figure 3.6 shows the two uses: example 60a shows the applicative use, while example 60b shows the retroapplicative. Note that the adjunct *that* in example 60b brings the patientive actant value brought by *water* in example 60a; moreover, *that* can be a trace, so there can be more than one lexeme in the same phrase carrying on the patient, as in example 60b.

As shown by Taylor [2002, 454], sometimes predicative uses are ungrammatical: *an utter fool* is perfectly acceptable in English, while in contrast **a fool is utter* is not. Hence, the predicative use can be blocked, according to the lexeme involved: this block in the lexical entry *utter* in the dictionary (see below).

Finally, there are some cases where it is simply impossible to determine if the adjunct is used attributively or predicatively.

- (61ab-it.) Paolo ha trovato il telefono rotto.

Figure 3.7 show two different construals belonging to example 61ab, which are equally valid: the left one shows the attributive use, the right one the predicative use.

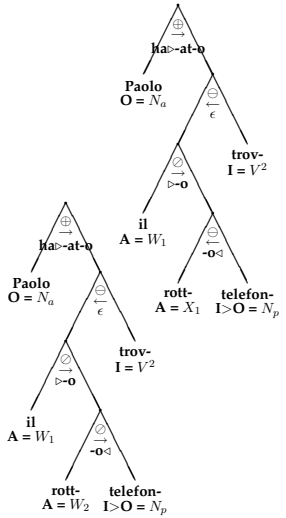


Figure 3.7: Adtrees of *Paolo ha trovato il telefono rotto* (61ab).

- (61a-en.) Paul has found the broken telephone.
- (61b-en.) Paul has found the telephone out of order.

In the construal 61a, Paul was looking for the broken telephone, while in the construal 61b, perhaps Paul was trying to telephone, but unfortunately the telephone was out of order. English clearly distinguishes the two construals with a suppletive strategy: *broken* for the attributive use, *out of order* for the predicative one.[14] However, it is noteworthy that also in English collocation plays a role; in fact, **Paul has found the out of order telephone* and **Paul has found the telephone broken* are both ungrammatical. In Italian, the lexeme *rott-* in example 61b works analogously to the example 38 already shown in Figure 2.33 (right): a whole relative clause is subsumed under a single lexeme.

The Italian example 61ab would generate two adtrees instead of one, i.e., a **family of adtrees**. This approach is correct, because this kind of ambiguity cannot be solved within the language system, but only with the help of pragmatic information not carved in the sentence. In other words, encyclopædic knowledge is involved here, and therefore these resolutions are definitely out of the scope of adgrams.

3.4 The dictionary

Adpositions are the scaffolds or meaning, while the lexicon gives the anchorage to the world model. How do they relate one to the other? Moreover, the lexicon is made not only by lexemes but also by affixes and some adpositions can be involved in transference. For instance, in example 58 (Figure 3.4) the adpositions *-ing* and *-able* are the morphemes which mark respectively the stativation of *read* and the adjunctivation of *enjoy*. Both morphemes *read* and *enjoy* are free, while in contrast the two adpositions are not; nonetheless, some adpositions can be free morphemes too, like for example the English preposition *across*.

What we need is the correspondence between morphemic freedom, adpositions and the lexicon, in order to give the inner structure of the dictionary as a whole. It is not possible to describe every class of morpheme in NLs such as English or Italian within a doctoral dissertation. Nonetheless, there are some general classes of morphemes, i.e., language-indipendent, that are necessary to build the dictionary. Here I am presenting these general classes while an extensive case study – namely, Esperanto – will be provided in the second part of this dissertation.

3.4.1 The double face of derivation

Some derivational morphemes are adpositions, while others are lexemes. In example 54, the English *-ly* and the Italian *-mente* shows that the set of ad-

Table 3.10: The morphemic classes in the dictionary

long class name	*short class name*	*grammar characterisation*	*morphemic freedom*
proper adposition	adposition	$\bot$	$\top\bot$
proper lexeme	lexeme	$\top$, any	$\top$
derivational adposition	transferer	$\top$, any	$\top\bot$
derivational lexeme	transferee	$\top$, any	$\top\bot$

positions comprehend not only prepositions and the like, but also a subset of what is traditionally known as derivational morphology. A derivational adposition is still an adposition as it does not have a grammar character but nonetheless it is involved in transference.

> **derivational adposition**: *a particular adposition which does not have any grammar character per se but it transfers the grammar character of the lexeme(s) it is applied to.*

I call standard adpositions **proper adpositions** in contrast, when precisation is needed. Derivational adpositions are also called **transferers**.

Adpositions can be free or bound morphemes; in this last case they are affixes. The vice versa is not necessarily true. In fact, not every affix is an adposition; some affixes do not participate in the process of the construction of the adtree. Hence, they are lexemes. Nor they participate in transference: their only function is to add semantic moduli to the applied lexeme. I call these morphemes derivational lexemes.

> **derivational lexeme**: *a particular lexeme which is applied to a proper lexeme only under certain grammar character restrictions.*

Derivational lexemes stand alone in an adtree leaf. They perform a derivation through the adding of semantics, without transference, i.e., without touching the grammar character of the lexeme(s) that govern that adtree leaf. I call derivational lexemes **transferees**. Transferees can be applied only at certain conditions, i.e., their application can be restricted to some grammar characters.[15]

The bricks of the language building are now complete. Table 3.10 shows the general structure of the dictionary.

Table 3.11: Homophones and equivalents of the English morpheme *in*

morpheme	*class*	*character*	*freedom*	*meaning*
in	adposition	$\bot$	$\top$	$\oplus$,[+SPACE|+TIME]
in	lexeme	x >E	$\top$	$\ominus$,[+SPACE,MOVEMENT]
in-	transferee	A< y	$\bot$	A,[+ANTONYM]
il-	transferee	A< y	$\bot$	A,[+ANTONYM]
im-	transferee	A< y	$\bot$	A,[+ANTONYM]
ir-	transferee	A< y	$\bot$	A,[+ANTONYM]
in-	transferee	O< y	$\bot$	O,[−PRESENCE]
il-	transferee	O< y	$\bot$	O,[−PRESENCE]
im-	transferee	O< y	$\bot$	O,[−PRESENCE]
ir-	transferee	O< y	$\bot$	O,[−PRESENCE]
un-	transferee	O< y	$\bot$	O,[−PRESENCE]

3.4.2 Transferees and homophony

Let me explain how transferees work through an example of a very small fragment of English. Some prepositions can be free or bound, like for instance the highly productive *in*, which is present in a lot of Romance languages, as well as NLs come into contact with Latin, such as English or German. In such cases, each use should be considered as a different modulus: these moduli have homophonous signatures. Moreover, the single modulus can have different signatures: *in-* has some equivalent signatures essentially for phonological reasons. Table 3.11 shows how the English morpheme *in* is represented in the dictionary. First, the difference between free and bound morpheme is taken into account; second, the right class has been found. The adposition *in* has been presented through examples 31,32,47 in chapter 2.

The lexical use is at the opposite spectrum of grammaticalisation, compared to the adpositional one. In the case of *in*, you can find it lexically in sentences like *come in*, *Liza and Paul were locked in*, *the car got in*. All these examples show that a modulus [+MOVEMENT] was activated, which inherits the modulus SPACE.

The adjectival transferee *in-* is bound and it has a lot of equivalent signatures, depending on phonology: for example, **in-liberal* becomes *illiberal*, and **in-possible* becomes *impossible*. In all these examples, *in-* has been applied to adjectives, a subset of the adjunctive class, and the semantic transformation is antonymous, i.e. the applied modulus is [+ANTONYM]. In fact, *illiberal* is the opposite of *liberal*, and *impossible* is the opposite of *possible*, and so on. When different affixes are equivalent signatures of the same modulus, they are called **allomorphs**.

In constrast, the stative transferee *in-* denotes a lack of presence: *in-advertence* is a lack of advertence, *in-appreciation* is the lack of appreciation. Note that it is not an antonymiser. In fact, *un-truth* is a *lack* of truth, not the opposite of truth, i.e., falsity.[16] Interestingly, in this case there is an additional equivalent signature, i.e., *un-*, that breaks the symmetry between the allomorphs of the adjectival and the stative transferees *in-*.

The variable y indicates the *result* of the transference, i.e., the final characterisation, while the variable x indicates the *beginning* of the transference. They have been used in Table 3.9 yet. For example, the stative transferee *in-* shown in Table 3.11 would be marked as O< y in an adtree description, y being the set of moduli brought by the morpheme, and O the final value of x after transference.

3.4.3 Transferers in action

An example of transferer in English is the morpheme *-ly*, which transforms an adjunct in a circumstantial, i.e., A>E. Now, let me show the conventions how transferers solve transference through example 65.

- (62.) Liza speaks informally.

The description of adtree leaves is now complete until the last detail. Only the semantic moduli, like for instance [+ANIMATE] for *Liza*, for readability and conciseness: readers can append them below the respective adtree leaves.

The word *informally* is made by four morphemes: *in-form-al-ly*. The lexical base is *form*, which is stative. This is adjectified by the transferer *-al* with a Slash adposition. The antonymiser transferee *in-* can be applied only now – resulting in *informal*. A proof of this is the fact that **in-form* does not exist, as it collides with the selective verb *to inform* (see also below). Finally, the final transferer *-ly* has been applied, so the final word *informally* has a circumstantial character, instead of being an adjunct, resolving the transfer chain O>A>E. The adtree of example 62 shown in Figure 3.8 depicts the process just explained. It is noteworthy that here grammaticalisation is at its maximum level, so all adpositions are retroapplicative and non-dimensional (Slash). As said before, actants can be attached only to trajectors, i.e., selective nouns or verbs, collocational O-lexemes or I-lexemes, or finally statified or verbified lexemes: that is why X_1 is attached to *form*.

Lexemes are the lexical heads of the adtree, and they are all free morphemes. They always are the rightmost participant in the current adtree, and the actant is placed there. Of course, a single node can be an adtree as well: in example 58, *read* and *enjoy* are both heads of respectively a stative subtree and a adjunct subtree. I will come back over this topic in section 3.5.

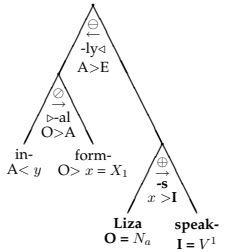


Figure 3.8: The adtree of *Liza speaks informally* (62).

In the sequel, I would give some examples in English of dictionary entries according to the fundamental grammar character of the basic morpheme. In fact, unlike other approaches, in adgrams each dictionary entry gives all the conventions for transfers, including suppletive ones – see again Table 3.7, if needed.

3.4.4 Verbal lexemes

The verbal characterisation is the most rich and important in the dictionary, because it gives the basic structure of each phrase. In particular, actants are individuated according to valence and the lexical entry. For this reason, a whole section in each lexeme is devoted to this crucial aspect. Table 3.12 shows how is treated the lexical entry *inform*. First, it is verbal in character: transfers are ordered according to grammaticalisation degree by convention, even if the transference to circumstantial should pass through stativation and adjunctivation before, as in this case.

Second, the verbal form *to inform* refers to the rules of variation according to parameters like diathesis, time, aspect, tense, gender and number. The analysis of words like *inform-ation-al-ly* refers to the adtree needed for this transfer. As already said in the introduction, adgrams describe how to build a grammar cognitive and computationally sound as if there would not exist any storage strategy in the mind. Of course, this is not true: it is very unlikely that an English native should build each time ex nihilo the adtree of frequent words like *informally* (rule/list fallacy); most probably, there is an adtree pattern such as *in-*α*-al-ly* stored and quickly activated, as only the lexeme α should be filled, while the adtree structure is already known – see again Figure 3.8.

It is noteworthy the different strategies for valence reduction and augmentation. The lexeme *inform* can reduce either invoking the paradigm behind the morpheme *oneself*, which is another lexical entry in the dictionary, or it can follow a complete suppletive strategy, invoking the verbal

Table 3.12: *inform* in the English adpositional grammar dictionary

verbif.	*circum.*	*adjunct.*	*stat.*	
I	I>E	I>A (⊖|⊘)	I>O	
to inform	inform-ation-al-ly	inform-ation-al	inform-ation	
moduli				
[+INFORM]	[+INFORM]	[+INFORM]	[+INFORM, +ABSTRACT]	
valence	1^{st}	2^{nd}	3^{rd}	*extra*
2	N_a	N_p	$\perp$	$\top\perp$
$2-1=1\rightsquigarrow$	to inquire			
REFLEXIVE $\rightarrow 2$	N_a	$N_p \rightsquigarrow$ oneself	$\perp$	$\top\perp$
$2+1=3\mapsto$	to cause			
	to inform			
	$N_a \rightarrow N_e$	N_a	N_p	$\top\perp$
PASSIVE $\rightarrow 1$	N_p	$\perp$	$\perp$	$\top\perp$

instance of *inquire* – the arrow $\rightsquigarrow$ means 'leads to' and it is a pointer to another lexical entry in the dictionary. In contrast, the augmentation invokes the pattern *to make* α as a mapping ($\mapsto$ means 'maps to'): valence values are shifted, and the Experiencer gets the first valence of this kind of complex structure. An analogous phenomenon happens when verbal diatheses are applied, with the moduli REFLEXIVE and PASSIVE. Note that reflexion does not hurt the valence value, but the referentiality of the actant, i.e., in the encyclopædia they are the Agent is also Patient, although in the syntagmatic axis Patientivity is carried out by the paradigm activated by *oneself*. The valence column 'extra' indicates if other actants, typically the Instrument, can be expressed or not within the same phrase in arguments external to verbal valences. In adtrees, extra actants are printed in plain, while valence actants are printed in bold.

Table 3.13 is similar. Note that the proper lexeme is *magnif* in the dictionary, while *-y* is only and adposition. Let us turn our attention to valence. The meaning of *magnif* in valence reduction is changed, even if the form did not: in terms of adgrams, V^2(magnif) and V^1(magnif) are homonymic. For example, in the sentence *these binoculars magnify well* the subject S is Instrument, while in *the risk magnifies the opportunities* the subject S is Patient, both having magnify as a monovalent verb. In contrast, in sentences like *Paul magnifies Liza's virtues* the verb is bivalent, the subject S is an Agent, while the object O is an Instrument, being *Liza* the Patient.

The choice between Instrument and Agent is made according to the modulus $\pm$ANIMATE brought by the O-lexeme in charge. The augmenta-

Table 3.13: *magnify* in the English adpositional grammar dictionary

verbif.	*circum.*	*adjunct.*	*stat.*	
I	I>E	I>A $\{\ominus\|\oslash\}$	I>O	
to magnif-y	magnif-ic-al-ly	magnif-ic	magnif-ic-ence	
moduli				
[+MAGNIF]	[+MAGNIF]	[+MAGNIF]	[+MAGNIF +ABSTRACT]	
valence	*1st*	*2nd*	*3rd*	*extra*
2	$\{N_x\|N_a\}$	N_p	$\perp$	$\top\perp$
$2-1=1$	$\{N_x\|N_p\}$	$\perp$	$\perp$	$\top\perp$
$2+1=3\mapsto$	to cause to magnify			
	$\{N_a\|N_x\}\rightarrow N_e$	$\{N_a\|N_x\}$	N_p	$\top\perp$
PASSIVE $\rightarrow 1$	N_p	$\perp$	$\perp$	$\top\perp$

tion is similar to *inform*, the case seen above.

Table 3.14 shows the entry of the lexeme *sink*. Adgrams treats NLs synchronically, not diachronically: even if the attributive adjective *sunken* derives from the past participle *sunk* of the verb *to sink*, it has its own entry in the dictionary (see below for details).[17] In this case valence reduction is nonsensical, while augmentation cause another shift of the actants' values, as shown in example 63. Each variant shows different levels of detail concerning the same construal.

- (63a.) N_p(The ship) sank.
- (63b.) N_x(The torpedo) caused N_p(the ship) to sink.
- (63c.) N_a(The admiral) caused N_p(the ship) to sink N_x(with a torpedo).

How to treat lexemes rich in idiomatic forms? Table 3.15 shows the basic forms of the lexeme *fall*. As already shown in section 2.9.1, this lexeme changes meaning according to the preposition which specifies the idiom, like in *Liza fell in love with Paul* and *Paul fell about Liza's imitations of the teacher*, therefore it will have an entry for each form. Let me take a rather extreme example of phrasal verb, *to fall in with*, exemplified by example 64.

- (64.) N_a(The lecturer) caused N_p(Liza) to fall in N_x(with the political views) of Kant.

Table 3.16 shows how *to fall in with* is represented in adgrams. Note that this phrasal verb is bivalent, unlike most phrasal verb with *fall* as the lexical head. Of course, in this case both *in* and *with* are false adpositions.

Table 3.14: *sink* in the English adpositional grammar dictionary

verbif.	*circum.*	*adjunct.*	*stat.*	
I	I>E	I>A	I>O	
to sink	$\perp$	$\rightsquigarrow$ sunken	sink-ing	
moduli				
[+SINK]	$\perp$	$\rightsquigarrow$ sunken	[+SINK, +ACTION]	
valence	*1st*	*2nd*	*3rd*	*extra*
1	N_p	$\perp$	$\perp$	$\top\perp$
$1+1=2\mapsto$	to cause to sink			
	$N_p\rightarrow\{N_a\|N_x\}$	N_p	$\perp$	$\top\perp$

Table 3.15: *fall* in the English adpositional grammar dictionary

verbif.	*circum.*	*adjunct.*	*stat.*	
I	I>E	I>A $\oslash$	I>O	
to fall	$\perp$	of the fall	fall	
		of the fall-ing	fall-ing	
moduli				
[+FALL]	$\perp$	[+FALL]	[+FALL, +ACTION]	
valence	*1st*	*2nd*	*3rd*	*extra*
1	N_p	$\perp$	$\perp$	$\top\perp$
$1+1=2\mapsto$	to cause to fall			
	$N_p\rightarrow\{N_a\|N_x\}$	N_p	$\perp$	$\top\perp$

Table 3.16: *fall in with* in the English adpositional grammar dictionary

verbif. I	*circum.* I>E	*adjunct.* I>A	*stat.* I>O	
to fall in with	$\perp$ $\perp$	of the fall-ing in with	fall-ing in with	
moduli [+FALL]	$\perp$	[+FALL]	[+FALL, +CHANCE, +INVOLVE]	
valence	*1st*	*2nd*	*3rd*	*extra*
2	N_p	$\perp$	$\perp$	with $\{N_a\|N_x\}$
$2+1=3 \mapsto$	to cause to fall in with $N_p \rightarrow \{N_a\|N_x\}$	N_p	$\{N_x\|N_a\}$	$\perp$

Table 3.17: *give* in the English adpositional grammar dictionary

verbif. I	*circum.* I>E	*adjunct.* I>A	*stat.* I>O	
to give	$\perp$	giv-ing	$\rightsquigarrow$ gift, $\rightsquigarrow$ present,...	
moduli [+GIVE]	$\perp$	[+GIVE, +BESTOW]	$\rightsquigarrow$ gift, $\rightsquigarrow$ present,...	
valence	*1st*	*2nd*	*3rd*	*extra*
3	N_a	N_p	$\perp$	to N_e
PASSIVE$_1 \rightarrow 3$	N_p	$\perp$	by N_a, to N_e	
PASSIVE$_2 \rightarrow 3$	N_e	N_a	$\perp$	by N_p

Table 3.18: *quite* in the English adpositional grammar dictionary

verbif. E>I quite	*circum.* E quite	*adjunct.* E>A quite	*stat.* A>O $\perp$	
moduli [+QUITE, +AGREE],	[+QUITE]	[+QUITE]	$\perp$	
valence	*1st*	*2nd*	*3rd*	*extra*
$\perp$	$\perp$	$\perp$	$\perp$	$\perp$

Table 3.17 shows the entry of the lexeme *give*, when not used as a phrasal verb, i.e., without false adpositions. Note that the adposition *-ing* is different from the one seen previously, e.g., *read-ing*, as the following test can clarify: *Paul is giving and supportive*. For stativation, the case of *give : gift* is analogous to the case of *sink : sunken* seen before. Suppletive strategies are always possible, as in the case of *present*. Example 28 – whose adtrees have been depicted in figure 2.23 – has already shown the two different passive diathesis transformations.

It could be a good exercise to write down the table of a phrasal verb like *give away*. This is left to the reader.

3.4.5 Circumstantial lexemes

Circumstantial lexemes are very rare, and most of them are circumstances, i.e., they do perform some grammar character transfer, usually along the adjunctivation line (only attributive, i.e., Slash). Some examples are: *quite*, *there*, *seldom*. Let me show a couple of examples: *quite*, which is a circumstance, and *often*, which is an adverb, i.e., it does not perform transference at all.

Table 3.18 shows how *quite* is represented in the dictionary. In sentences like *I quite like it* it modifies the verb, hence it acts as a circumstantial, while in sentences like *She is quite young* it modifies the adjective *young*, hence it acts as an adjunct. At least in Britain, *quite* can also be used as a single word sentence, to express a complete agreement, like in example 65: this is a verbified use.

- Liza: I don't want to talk about him now!
- (65.) Bea: Quite (so)!

In terms of adtrees, example 65 would be represented through mirroring, as a marked syntactic construction, like presentative verbs – see Figure 3.9.

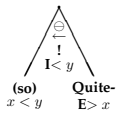


Figure 3.9: The adtree of *Quite (so)!* (65).

Table 3.19: *often* in the English adpositional grammar dictionary

verbif.	*circum.*	*adjunct.*	*stat.*	
E>I	E	E>A	A>O	
⊥	often	⊥	⊥	
moduli				
⊥	[+OFTEN]	⊥	⊥	
valence	*1st*	*2nd*	*3rd*	*extra*
⊥	⊥	⊥	⊥	⊥

The case of *so* is very interesting. For instance, in sentences like *I think so* it modifies the verb, hence it is E in character; however, most often it is A in character. The interesting part is that it can be applied at whatever grammar character: in sentences like *I feel so good* it modifies the circumstantial; in sentences like *I am so jealous* it modifies the adjunct; last, in the idiomatic expression *so what?* it modifies a stative placeholder. In all cases, *so* is a derivational transferee, i.e., it adds semantic moduli to what is applied, without changing the grammar character.

Table 3.19 shows the very simple table of a circumstantial by selection, i.e., the adverb *often*. It is quite unusual to find morphological paradigms within circumstantial lexemes. An example can be the series formed by *-ward(s)*: *to-ward(-s)*, *on-ward(-s)*, *for-ward(-s)*, *back-ward(-s)*, *right-ward(-s)*, *left-ward(-s)*, *up-ward(-s)*, *down-ward(-s)*, etc. Series like this should be listed apart, analogously to the case of *to fall in with*. Table 3.20 shows the case of *forward*. The modulus FORWARD is an hyponym of the modulus SPACE by definition, so it can be omitted. It is noteworthy that two equivalent signature exist for the E-lexeme, i.e., *forwards* and *forwardly*, and three different stativations. As a noun, *forwards* can be used in SPORT, like basket, hockey, or football, meaning an attacking player, or in finance, meaning *forward contract*. These two special meanings would be activated if the moduli are activated within the phrase – see chapter 8 for details. If the abstractive derivational adposition *-ness* is applied, the third meaning will be automatically activated.

Table 3.20: *forward* in the English adpositional grammar dictionary

verbif.	*circum.*	*adjunct.*	*stat.*	
E>I	E	E>A	A>O	
⊥	for-ward(-s)	for-ward(-s)	for-ward-s$_{a,b}$	
	for-ward-ly		for-ward-ness$_c$	
moduli				
⊥	+[FORWARD]	+[FORWARD]	+[FORWARD]$_{a,b,c}$	
			+[SPORT]$_a$	
			+[FINANCE]$_b$	
			+[ABSTRACT]$_c$	
valence	*1st*	*2nd*	*3rd*	*extra*
⊥	⊥	⊥	⊥	⊥

Table 3.21: *elephant* in the English adpositional grammar dictionary

verbif.	*circum.*	*adjunct.*	*stat.*	
O>I	O>E	O>A	O	
to be				
an elephant	?elephant-ine-ly	elephant-ine	elephant	
moduli				
[+ELEPHANT, +BE]	[+ELEPHANT]	[+ELEPHANT]	[+ELEPHANT]	
valence	*1st*	*2nd*	*3rd*	*extra*
BE $\rightarrow$ 2	N_p	N_a	⊥	⊥

3.4.6 Stative lexemes

In the English language there are some nouns, i.e., stative lexemes by selection, as seen previously in section 3.1. Table 3.21 shows the entry *elephant*. Example 59a already explained how actants work with the proper verb *to be* – see again Figure 3.5.

Some stative lexemes, like *day*, apparently do not perform verbal transference. Table 3.22 shows the dictionary entry of *day*. It is noteworthy that the circumstantiation is more prominent than adjunctivation. The case of *book*, shown in Table 3.23, is more complex. First, the verbification does not even have the modulus BOOK active. The moduli RESERV and ACCOMOD are written along the corrispondent dictionary entries: e.g., *accomod-ate : accomod-ation*, *reserv-e : reserv-ation*.

Table 3.22: *day* in the English adpositional grammar dictionary

verbif.	*circum.*	*adjunct.*	*stat.*	
O>I	O>E	O>A	O	
⊥	to-day	to-day's	day	
moduli				
⊥	[+DAY]	[+DAY]	[+DAY]	
valence	*1st*	*2nd*	*3rd*	*extra*
⊥	⊥	⊥	⊥	⊥

Table 3.23: *book* in the English adpositional grammar dictionary

verbif.	*circum.*	*adjunct.*	*stat.*	
O>I	O>E	O>A	O	
to book	⊥	book-able	book_a	
			book_b	
			book-s_c	
			Book_d	
moduli				
[+RESERV]	⊥	[+BOOK,	$[+\text{BOOK}]_{a,b,c,d}$	
[+ACCOMOD]	⊥	+POSSIBLE]	$[+\text{TELEPHONE}]_b$	
			[+RECORD,	
			$[+\text{SET},]_c$	
			$[+\text{BIBLE}]_d$	
valence	*1st*	*2nd*	*3rd*	*extra*
3	N_a	N_p	⊥	for N_e
$\text{PASSIVE}_1 \rightarrow 3$	N_p	⊥	⊥	by N_a, to N_e
$\text{PASSIVE}_2 \rightarrow 3$	N_e	N_a	⊥	by N_p

- (66a.) N_a(Bea) has booked N_e(them) N_p(a table) at the restaurant.
- (66b.) N_p(A table) was booked N_e(for them) N_a(by Bea).

Lexicalisation and stative compounds

What to do in words with two stative lexemes, like *book-mak-er*? Should they fall under the rubric of *book* or under the rubric of *mak(e)*? In reality, this words are the result of a heavy grammaticalisation occurred diachronically, as the meaning became more and more specific and fixed: *the maker of books → the maker of books → the bookmaker*. I call this particular grammaticalisation **lexicalisation**. As English is morphologically a German language, the head is the second stative lexeme, while the tail is the first one. The adtree will be built accordingly, as shown in example 67.

- (67.) The bookmaker paid out the winning.

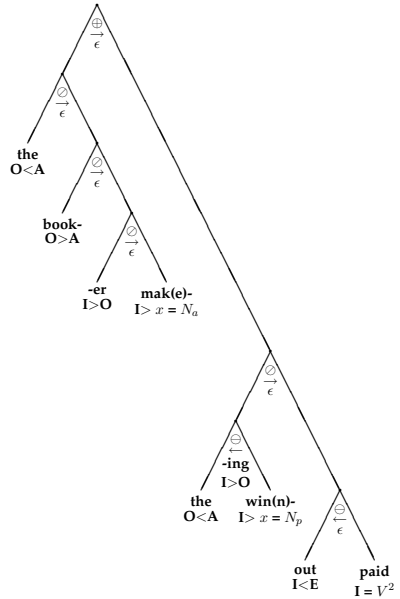


Figure 3.10: The adtree of *The bookmaker paid out the winning* (67).

Table 3.24: *nice* in the English adpositional grammar dictionary

verbif.	*circum.*	*adjunct.*	*stat.*	
A>I	A>E	A	A>O	
⊥	⊥	nice	⊥	
moduli				
⊥	⊥	[+NICE]	⊥	
valence	1^{st}	2^{nd}	3^{rd}	*extra*
⊥	⊥	⊥	⊥	⊥

Figure 3.10 shows the adtree of example 67. The meaning of brackets in lexemes like *win(n)-* or *mak(e)* is due to a phonological variation of English in morphological composition: this will be rendered with ad hoc, very specific rules. The phrasal verb *to pay out* has an additional internal argument (*out*) which does not carry any actant, and in fact it is fixed in advance. The adtree is complex, compared to the number of words involved. Most probably, the rule/list fallacy is valid here too: a paradigm of stative compounds governed by *-mak-er* is certainly active, with common words like *shoe-mak-er* and *cabinet-mak-er*.

Finally, note that foreign compounds borrowed in English should not be decomposed, as the average speaker is totally unaware of its structure, and in fact they are no longer productive. For instance, *blitzkrieg* and *kriegspiel* are single lexemes in English, while in German they retain their productivity: *Blitz-krieg*, *Blitz-aktion* (literally, 'flash-action'), *Blitz-schlag* (literally, "flash-lightning", i.e., 'love at first sight'), and also *Krieg-spiel*, *Krieg-s-film* (war film), *Krieg-s-ge-fang-en-e* (prisoner of war), etc.

3.4.7 Adjunctive lexemes

Adjuncts comprehend all adjectives, both attributive and predicative (see again section 3.3.4), and a lot of so-called 'functional words', like determiners (*this, these*) and articles (*a, the*). Analogously to the other types, some adjuncts are selective, while others are not. For example, *nice* is a selective adjective (Table 3.24), while *blue* is collocational (Table 3.25). In fact, in sentences like *I feel blue* the morpheme *blue* modifies the verb *feel*, hence it is circumstantial, while in *the blue sky is deeper in Norway than in Italy* the morpheme *blue* is clearly an attributive adjective. Note that the stative use of *blue* coincide with the explicitly abstractive *blue-ness*.

Table 3.25: *blue* in the English adpositional grammar dictionary

verbif.	*circum.*	*adjunct.*	*stat.*	
A>I	A>E	A	A>O	
to blue	blue	blue	blue	
			blue-ness	
moduli				
[+BLUE, +BECOM]	[+BLUE]	[+BLUE]	[+BLUE, +ABSTRACT]	
valence	1^{st}	2^{nd}	3^{rd}	*extra*
2	N_a	N_p	⊥	⊥
$2-1=1 \mapsto$	to be blue			
	N_p	⊥	⊥	⊥

3.5 The notion of grouping

I postponed to notion of grouping to the end of the first part as readers should be familiar to adtrees, adtypes and actants before to approach this concept. In fact, the theoretical basis of groups is far different from the one proposed by the Chomskyan tradition, although the very concept of grouping is due to Chomsky himself. This is the definition of group:

> **group**: *a group is formed by one or more adtrees, i.e., a set of syntagms, where a syntagm is a triple lexeme-adposition-lexeme. Its head is always lexical, and it coincides with the lexeme bringing the actant value, either full or void. Groups can be nested one into the others.*

In adgrams, the concept of grouping is complementary of the concept of adtrees: while adtrees shows the adpositional structure behind lexemes, transferers and transferees, groups are a way to put together adtree parts in semantically coherent parts, like what is usually done in the Bloomfieldian immediate constituent analysis (perhaps in a more readable way, as there is no so heavy bracketing).[18] Eventually, the concept of group concides with the notion of syntagm.

Hereafter ther are some rather complex examples of groups, taken from already analysed examples: the most interested reader can retrieve the adtrees and compare the two notations. It can be useful also to check again section 2.8.

- (27.) N_a(Somebody) V(called) N_p(me) X_1(at the phone) Z_1(with a mobile).
- (31.) N_p(Carlo) V(cade) N_x(nel tombino).

- (32.) N_p(Carlo) V(fuma) Y_1(nel tombino).
- (33.) N_p(Liza) V(is) N_a(blonde).
- (39.) X_1(Domani) V(piove).
- (42.) N_a(Al) V(studies) Y_1(in the library) $\otimes$(and) N_a(Carl) V(plays) X_1(home).
- (43.) N_a(Alfred) V(peut X_1(payer)), $\oslash$(parce que) N_p(il) V(est) N_a(riche).
- (45.) V(N_p(Si) scatenò) N_a(un finimondo) $\otimes$(:) Z_1(urla) $\otimes$(,) Z_2(insulti) $\otimes$(,) Z_3(scherni) $\otimes$(e) Z_4(ingiurie).
- (46.) N_p(The library $\oslash$(where N_a(I) X_1(often) V(study) N_p(literature))) V(is) N_a(W_1(far) away).
- (54.) N_p(The restaurant) V(is opened) X_1(Monday X_2(evening)).
- (55.) X_1(Suddenly) N_a(Liza) V(calmed) N_p(W_1(her) walk).

Especially if the final adtree is very complex, grouping is a compact notation to enlighten some phenomena over others, compared to adtrees. It is noteworthy that a each group is equipollent to (a set of) adtree(s).

Notes

[1] Here I am paraphrasing Quine [1987, 130] as cited by Taylor [2002, 63].

[2] The article referred as Whorf [1945] is posthumous, but it was written in 1937. At the time it was out, the linguistic mainstream wind in America was changing, so no one paid attention to it. I found a relevant use in Miner [2008]. I use examples by Whorf in NLs I have some knowledge, i.e., I will not refer to Hopi or Yana. However, this fact does not prejudge the line of reasoning presented hereafter.

[3] The other two examples by Whorf *longevity* and *altruism* offered by Whorf are not convincing as selective nouns. In fact, applying morphemic analysis, we can easily say that English words that end in *-ity* and *-ism* are always nouns. More precisely, these two morphemes are signatures of the [STATIVATION] modulus, which also act as vehicles of semantic features. This error happened because Whorf based himself on words, instead of morphemes.

[4] The term 'stativation' for the polarity opposed to 'verbification' is unusual but appropriate, as it does not suggest derivation as 'nomination' or 'nominalisation'. Moduli at the same hierarchical level are not derivation one from the other.

[5] The phenomenon exemplified by 53c is sometimes called 'ad sensum agreement'. In other cases, both possibilites are grammatical, as in the sentence *for this receipe two cups of sugar* {*is*|*are*} *needed*. Perhaps covertion happens because diachronically moduli have lost productivity as paradigms. For examples think at the so-called 'strong verbs' in English: *run ; run ; run* and *put ; put ; put* are part of a paradigm which has lost its productivity in favour of the paradigm α *;* α*ed ;* α*ed* where α is a 'weak verbal root'.

[6] Perhaps, a subtle difference of register can be found, being *don't* less informal than *do not*. However, this kind of differences pertains the communicative purposes of the speaker, i.e., pragmatics, as well as phonological attraction phenomena (e.g., in English we say *black and white* instead of *?white and black*, while in Italian it is exactly the opposite: *bianco e nero*) As explained in the introduction, these nuances do not enter the adpositional grammar model for now.

[7] We can follow a similar line of reasoning for the three signatures *-are*, *-ere*, *-ire* for the modulus [INFINITIVE] in Italian, which is active only when the modulus VERBIFICATION is active. Many examples are possible.

[8] I prefer 'adjunctivation' instead of 'adjectivation', the term originally used by Whorf, as the phenomenon which is referring to is broader than adjectives.

[9] A *circumstance* is a lexeme with the E-modulus, while a *circumstantial* is the node of the adtree where the circumstance is applied.

[10] This translation was proposed by Lucien Tesnière himself [Tesnière, 1959].

[11] Truly, the lexeme *muggl-* is understable by a Harry Potter's fan, like me.

[12] I will come back into the topic of traces in the second and the third parts of this dissertation.

[13] The approach of having a limited set of deep cases or thematic roles which ontologically exit before NLs was postulated first in the 17^{th} century by John Wilkins as the *Real Character* [Eco, 1993]. This approach survives the centuries: a whole school of machine translation was found on the same principles, called 'interlingua' (see at least Arnold et al. [1994] and Hutchins and Somers [1992], for a survey). For example, in the Universal Network Language (UNL) project, a sentence like *Paul broke the window with the hammer* would be analysed as follows: *Paul* is agent, *broke* is cause, *the window* is benefactive and *with the hammer* is instrument. No syntactical analysis is even performed, and that is why it is called 'interlingua', because this representation does not depend on a single NL (for UNL, see at least Cardeñosa et al. [2005]). Although this approach seems promising, it does not use any regularity present in NLs, and hence it needs a huge amount of strong supervision in order to build the dictionary.

[14] In chapter 5 I will show the strategy of Esperanto in such cases with the same example.

[15] The terms 'transferer' and 'transferee' are not directly used by Tesnière [1959], but they can be derived from the Tesnerian concept of transference. In French, 'transferer' would be rendered as *translateur*, while 'transferee' would be translated as *transféré*, using two different lexemes: *transf-* and *transl-*. In contrast, Italian would use only one lexeme, *trasl-*: *traslatore* for 'transferer' and *traslato* for 'transferee'. A possible translation in Esperanto can be the couple *transigilo*, 'transferer', and *transigito*, 'transferee' (also see the Appendix).

[16] That is why it was marked as [−PRESENCE] instead of [+ABSENCE], which would have been wrong.

[17] The limits of the morphological analysis is synchronicity. It is of no use in adgrams to point out the Latin morphemes *-er* and *-ter* in words like *inf-er*, *sup-er*, *int-er*, *dex-ter*, *sinis-ter*, clearly made by two morphemes, and transmitted as fossil in a lot of actual NLs. Analogously, the entry of *utter* will not be put into relation with *out*, even if diachronically the two words are rilated: *utter* is a contraction of *out+er*.

[18] The immediate constituent analysis (ICA) was introduced in modern linguistics by Bloomfield in 1933, but its roots are at least in the Italian grammarians of cinquecento – see Percival [1976] in McCawley [1976].

Part II

Application

Chapter 4

The translation game

What 'machine translation' does exactly mean? First, I call **translation** a process of rendering a text, i.e., a coherent chain of grammatical sentences written in a given NL, from a source language ($\mathcal{L}_s$) into a *reliable* text written in a target language ($\mathcal{L}_t$). If $\mathcal{L}_s = L_t$ is true, the term **paraphrasis** will be used. 'Reliable' means that:

1. the text in $\mathcal{L}_t$ is syntactically correct on reading, i.e., without any further editing;

2. the original meaning in the $\mathcal{L}_s$ Spanish is preserved in the $\mathcal{L}_t$ version, i.e., both texts evoke the same content, even if the construals are entirely different.

Second, note that I use the restricted sense of 'translation', referring only to *asynchronous written texts*. I use the term 'interpretation' for synchronous oral discourse translation, as it is used in the current EU language policy literature [Phillipson, 2003]. It is noteworthy that interpretation is out of the scope of this dissertation.

For **machine translation (MT)** I simply mean a translation carried out by a computer, without any human aid. There are three main theoretical paradigms in MT research: the direct paradigm; the transfer paradigm; the interlingua paradigm. The **direct paradigm** is the oldest: the system is a series of bilingual dictionaries, and the translation is performed word-by-word. No linguistic analysis is performed. They are the less accurate. The **transfer paradigm** is a rule-based system: there is a set of rule for each couple of NLs and a translation arrow ($\mathcal{L}_s \Rightarrow \mathcal{L}_t$). A detailed morphosyntactic analysis is performed. Finally, there is the **interlingua paradigm**, where the machine performs a pragmatic and semantic analysis of the source language $\mathcal{L}_s$ – usually using a longer list of actants, and perhaps some form of case-based reasoning – then the same procedure is applied to the target language $\mathcal{L}_t$. No morphosyntactic analysis is performed.[1]

Until the 1990s, MT systems were top-down – i.e., linguistic corpora were annotated with information with supervision, by human informants – regardless of the theoretical approach. Then, existing linguistic corpora were starting to be used, often parallelised, either on a text or sentence level. Machine learning methods, usually based on statistics, are used nowadays. Interestingly, the first MT systems driven by corpora followed the direct paradigm, while since year 2000 all systems follow the transfer paradigm.

If translation is performed by a computer and then revised by a human being, the acronym **CAT (computer-assisted translation)** will be used. CAT includes the use of **translation memory (TM)** tools as a special case. A TM tool is a knowledge base where translation chunks are stored, so that when a human translator encounters an analogous chunk of text the system proposes the correspondent translation already approved. This approach proved to be highly effective in domains where a controlled, standardised language is adopted, such as technical manuals. TMs can be considered an explicit and technical answer to the rules/list fallacy argument.

Why should computational linguistics care about MT? This question is twofold, the first one being computational, the second one linguistic. They are logically indipendent one the other, so I will treat their answers separately.

The computational side of the answer is simpler than the linguistic one. For MT people generally mean the scenario presented in the next section, that I call the 'default MT scenario'. In this chapter I propose an alternative scenario, that I consider more appropriate to the linguistic purposes of adgrams and, more generally, for computational linguistics as a research area. I describe each scenario as a *Gedankenexperiment*, i.e., a thought experiment.

4.1 The default machine translation scenario

Alice is a native speaker of a $\mathcal{L}_s$, let it be Spanish, and she wants a written document of hers, e.g., a newspaper article, to be read by Bob, who does not understand anything in Spanish. Bob is a native speaker of the $\mathcal{L}_t$, let it be Tamil. The double arrow ($\Rightarrow$) indicates that a translation is performed; the input is the first element on the left, while the output is the second element on the right of the arrow. The greater ($>$) means "writes" while the lesser ($<$) means "reads". Inf formulae:

$$A > \mathcal{L}_s \Rightarrow B < \mathcal{L}_t$$

As it is neither simple nor cheap to find human professional translators from $\mathcal{L}_s$ to $\mathcal{L}_t$, Alice decides to put her text into a machine – i.e., a computer program – that is supposed to give a reliable translation in Tamil.

As so described, MT is a linguistic task that simply cannot be done non-computationally.

4.1.1 The problem of machine translation evaluation

The linguistic side of the answer is less trivial. I argue that MT should be *the* testbed of language models, i.e., instances of general language descriptions and formal linguistic theories, which aim to have at least a cross-linguistic validity – perhaps it is the only genuinely computational testbed we have. More clearly stated, a reliable MT is a proof of validity of the underlying model. But a major problem arises in this scenario in order to reach this goal: how to evaluate a MT task?

The most often used method in current research is to compare the MT output with a gold standard translation from which researchers extract features to train automatic metrics. However, as convincingly shown by Callison-Burch et al. [2006], automatic metrics such as BLUE are very useful to improve MT systems during their development but they may strongly differ from human evaluation and in any case they cannot replace them.

On the other hand, human evaluation is subject to a great variability which is not easy to control or measure. Furthermore, the fact that human evaluators know that the outputs are made by machines and not by human translators deeply influences, if not invalidates, their final evaluation. To solve this problem, which is of the greatest importance, I propose here an alternative scenario.

4.2 The translation game scenario

Suppose that Charles is the Spanish-Tamil MT designer, who enters the default scenario. Charles asks Alice not to write directly in her mother tongue but in a special controlled language, meaning not a domain control but a *fully controlled language system*, i.e., a **quasi-natural language (QNL)** as defined by Lyons [2006]. A QNL is non-natural without necessarily being unnatural, i.e., it is non-natural in a very different way compared to predicate calculus or computer programming languages such as Pascal, C, Prolog, Smalltalk or ML. A QNL shares all the basic properties of a NL, the most important being the double articulation, i.e., the divisions either in phonemes or in morphemes. Furthermore, a QNL has the following properties:

1. it is highly regular in morphology.
2. As a direct corollary, homophony and suppletion are less used lexical strategy compared with NLs.
3. Part-of-speech tagging is easier as there is almost no allomorphy.

4. Lexemes normally are highly productive, i.e., lexical paradigms as shown in chapter 3 normally do not have arbitrary blocks.

QNLs are not restricted in semantics, i.e., they have the same expressive power of NLs. It is noteworthy that both non-pathologic child languages and planned languages belong to this category.

> For example, QNL has a subclass Quasi-English, one of whose memebers is like English in all respects except that it is inflectionally regular, all plurals of nouns being formed with the *-s* suffix (*childs*, *sheeps*, *gooses*, etc.), all past-tense forms of verbs with *-ed* (*goed*, *runned*, *beed*, etc.), and so on. This is a language part of which children construct ofr themselves (an then in part decostruct – if I may so express it at) at a certain stage in the normal (natural_3) process of acquiring English. It is also the language into which English would presumably have developed under particular environmental conditions which maximized the effect of what is traditianally referred to as analogy. [Lyons, 2006, 69–70]

Coming back to our *Gedankenexperiment*, let assume this QNL being the Entry language ($\mathcal{L}_e$) of the MT system used by Alice. The $\mathcal{L}_e$ will generate the texts in both the $\mathcal{L}_s$ and in the $\mathcal{L}_t$ at the same time. The $\mathcal{L}_e$ should simplify not only Charles' work, but also Alice's. In fact, not only Alice's text in the $\mathcal{L}_e$ will be rendered simultaneously both in Spanish and in Tamil, but she can also compare her text with the Spanish MT and hence adjust it, learning how to improve her text and her use of the MT system simultaneously.

In particular, the system has a TM tool which proposes second or third choice alternative translations. In fact, when Alice judges the Spanish MT to be satisfactory, she will push an "ok, I'm satisfied" button. From that moment, the system will keep track of Spanish and Tamil texts as TMs – eventually adapting the generation rules according to the satisfactory version, so to have a better proposal of Spanish next time. In the background, the Tamil generated text will be adapted too, according to their respective generation rules. In formulae:

$$A > \mathcal{L}_e \Rightarrow A < \mathcal{L}_s \cap S > \mathcal{L}_t \cap B < \mathcal{L}_t$$

I argue that there are two compensation for Alice's additional work in writing in the $\mathcal{L}_e$ instead of in her native tongue Spanish, although I admit this is highly debatable. First, Alice will work *bilingually*: the comparison between the between the text in the $\mathcal{L}_e$ and the text in the $\mathcal{L}_s$ will improve the clarity and conciseness of her text.[2] Second, the system would be inproved by the increase of TMs and the adaptation of the generation rules from each language pair.

Adgrams should solve the problem of NL understanding, as it covers the model for morphosyntax and semantic description of the single NLs as

well as the rules for MT. However, it is noteworthy that adgrams work as if there is no TM involved. Following the classification of MT systems, the system proposed here is basically a rule-based transfer system [Hutchins and Somers, 1992]. An apart role is given to the $\mathcal{L}_e$, as it generates both texts in the $\mathcal{L}_s$ and $\mathcal{L}_t$ respectively, so the implementation of $\mathcal{L}_e$ should be more fine-grained. Moreover, the notion of grammar character can be interpreted as a sort of interlingual level.[3] Finally, the structure of the dictionary, as described in chapter 3 can be viewed as a collection of examples, i.e., "machine translation by example-guided inference, or machine translation by the analogy principle" [Nagao, 1984, 4].

A detailed description of the MT architecture following the translation game scenario is postponed to chapter 8, as an instance of the $\mathcal{L}_e$ is needed.

4.2.1 A new evaluation test

From Bob's point of view, nothing changes between the two scenarios. He may be totally unaware of all this process, all he needs is to be sure about Tamil text reliability – in the sense given above – of Alice's Spanish original. The main advantage in the translation game scenario is in the test we are going to perform, which is not possible in the default scenario.

Suppose now that Dave is a native bilingual speaker of $\mathcal{L}_s$ and $\mathcal{L}_t$: he can read both Spanish and Tamil texts with a reading mastery (C2) level, according to the classification used in the Common European Framework of Reference for Languages.[4] Dave is neither a computer scientist nor a linguist, but he is a professional translator. He knows neither Alice nor Bob and cannot access the text written in $\mathcal{L}_e$ by Alice.

Charles, the MT designer, asks to Dave if the text in Spanish and the text in Tamil are reliable translations one of each other. It is important that Charles does not specify the translation arrow – i.e., which is the $\mathcal{L}_s$ and which the $\mathcal{L}_t$ – nor the fact that Dave is evaluating *machine* system outputs. What really matters is the *reliability* of the texts. The key question is: are the texts in Spanish and Tamil a translation one of the other? To what extent?

This test leads to a neutral evaluation, without any prejudice due to the fact that a machine, not a human, had producted *both* texts. In fact, unlike the default scenario, Dave is evaluating two really comparable documents, instead of a human-produced document versus a machine-produced one.

I consider that the test would be passed if Dave will find the same amount of translationese – i.e., translation-specific linguistic signals that reveal a document being a (human or not) translation – in both texts. In the optimal case, Dave would not be able to decide whether the Spanish is the original version or not.

Finally, it is noteworthy that this test does not prejudice any use of automatic metrics: it is simply an additional test fulfilled instead of the default

human evaluation tests.

4.3 Esperanto as the entry language

[1]Esperanto is a QNL and a planned language which is best fit for being the $\mathcal{L}_e$ in the translation game. Here I explain the reasons behind this choice, based on some facts about Esperanto as a sociolinguistic phenomenon. Some reasons are theoretical, due to its unique linguistic status, and are investigated in the next subsection. Other reasons are practical, and they are presented in the following subsection. Finally, the full description of the adgram of Esperanto will be given in the next chapter.

4.3.1 The quest for the international auxilary language

Since the end of the 19^{th} century up to the first half of the 20^{th} century, the quest for a the definitive form of the **international auxiliary language (IAL)** was a hotly debated issue among linguistic scholars and amateurs. In fact, about 1,000 language projects were proposed in that period, especially in Europe: the strongest effort for cross-cultural communication ever made.[5]

An IAL is a QNL for a specific purpose, "oral and written use between people who cannot make themselves understood by means of their mother tongues", as clearly stated by Otto Jespersen in 1931. IALs are a special kind of **planned languages**: fully linguistic systems – *langue*, in Saussurean terms – launched by an author through a book (or a web site, in more recent times), containing grammar and the basic vocabulary.[6] When launched, a IAL is a *langue* but not a *parole*, i.e., there is no linguistic behaviour performed by the speech community members, simply because there is not any speech community – note that this definition excludes the so-called a priori languages as pasigraphies, such as John Wilkins' *Real Character*, or François Sudre's Solresol.[7] Therefore, the first step in the analysis of a planned language is to detect the very moment the language gets published. Crucial issues as standardisation, graphisation and reform become in fact very different after the **primitive contract** is settled, as stated by Ferdinand de Saussure:

> The primitive contract gets confused with everyday's language life. A system of signs as a language is received passively by the next generations. In any case the system of signs have the property to be transmitted by laws of their own, indipendent from the ones that settled the original contract (even if there is an explicit agreement, as in the case of Esperanto). The moment the contract is accepted, *no one is*

[1]Section 4.3 updates and summarises some publications of mine: Gobbo [2008a], Gobbo [2006] in Borbone et al. [2006], Gobbo [2005a], Gobbo [1998].

> *the owner anymore*. A language is like a goose brooded by a hen. After that moment, a language enters its semiological life, and it is impossible to get back (emphasis by me).[8]

The primitive contract involves tha fact that a (Q)NL cannot be re-planned – i.e., touched in the language core: phonetics, morphology, syntax – even if a single author is identifiable, as the (Q)NL creates its own mystique of appartenence and permanence while entering its "semiological life" – that is, when a speech community produces the *parole* (Figure 4.1). Most IALs

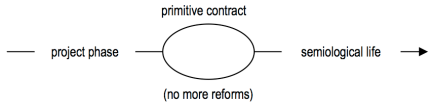


Figure 4.1: The life of a planned language [Gobbo, 2008a, 28]

never succeed to establish a speech community: Esperanto is the most notably exception.

4.3.2 Esperanto and its evolution

Esperanto was launched by Ludwik Lejzer Zamenhof (1859-1917), an Ashkenazi intellectual, follower of the Jewish Enlightment (Haskalah) and early Zionist. Moving from the debate about the possible solutions of the Jewish Question, Zamenhof spent his life along a twofold project: a 'neutral-human' language and a 'neutral-human' culture and nonclerical religion, that could serve as an enlighted bridge equally for Jews and non-Jews. The language should be the main tool for that non-ethnic culture.[9] Initially he turned his attention to Yiddish, and wrote the first grammar of that language ever; then, after the fall of the American option for Zionism, he turned his attention not to Jews only but to mankind: every people could take part this non-ethnic culture through the non-ethnic language. His ideals formed the so-called *interna ideo*, internal idea, the kernel of the esperantic philosophy.[10] He started in 1887 from the language, publishing a book in Russian, with the grammar, some literary texts (original and translated) and the basic dictionary. Thereafter, the same book was published in Polish, French, German, English and Swedish.[11] Zamenhof signed the book under the pseudonym 'Doktoro Esperanto', which eventually became the name of the new IAL. The language core of Esperanto – i.e., phonology and morphosyntax – is heavily influenced from Zamenhof's mother tongues, Yiddish and Belarussian, while the lexicon is made of Romance, Germanic and Slavic morphemes, in this frequency order (see chapter 5 for details).

At the end of the 19^{th} century there was a certain amount of groups and societies supporting the idea of a IAL, and they were unsatisfied by

the most popular IAL at that time, Volapük, so they turned their support to Esperanto.[12] Initially the language was used in written form: a lot of discussions were about grammar. Zamenhof followed a precise language policy: supporters were free of using the language for whatever goal (and that meant that new words should enter as neologism, as in NLs), while they should adhere strictly to the 16 basic grammar rules of the language, which should act as a barrier to language variability.[13] Zamenhof was true, as the primitive contract was established among the Esperanto community; nevertheless, people were proposing a lot of structural reforms (e.g., changing the way to form plurals). Therefore, in 1894, a referendum was imposed to the readers of the first journal, *La Esperantisto*, about the possibility of remaining loyal to the original system or to shift to a new one, which collected the most frequent remarks. The voting resulted favourable to the original system, and the possibility of structural reforms were declared out of the Esperanto community; this was declared again in the first congress occurred in Boulogne-sur-Mer (France) in 1905: since that moment, Esperanto became a full used language, also in speech, with a relevant public life. We can say that, since 1905, Esperanto became a QNL in Lyons' terms.

Of course, the debate about the definitive form of the IAL did not cease in 1905. A lot of offsprings, reforms, alternative projects, were launched after that date. Only a couple of them, i.e., Ido and Interlingua, are still in used nowadays.[14] However, no other IAL succeeded to develop at such a level as Esperanto: the community has succeeded to survive two world wars – in spite of the persecutions by Nazis and Stalinists.[15] Hence, Esperanto succeed to build a non-ethnic international culture of its own, with an apart attention to original literature [Sutton, 2008]. Realistic estimates are of 50,000 members of the Esperanto speech community, i.e., people that actually use the language in their life. In minimal part, Esperanto is also acquired by children in multilingual families. This is a known and studied linguistic phenomenon, sometimes compared with the revitalisation of Modern Hebrew. However, it is still debated if it is acquired really as a first language, as there is no truly monolingual speaker of Esperanto, and Esperanto is always in the weak position in bilinguism, as there is no *Sprachraum*, language area, belonging properly to that QNL.[16]

The speech community evolved and so the language: some forms became archaisms, while others are constantly invented, most notably in the field of computer science.[17] There are some idiomatic expressions, exactly as in NLs [Fiedler, 1999]. This means that Zamenhof's ideas, views, and use of Esperanto are still important to understand the kernel of Esperanto, but contemporary Esperanto is no more Zamenhof's language from a lot of time: in some cases the speech community made emerge linguistic phenomena that Zamemhof could not be aware of. Table 4.1 shows how the official dictionary of Esperanto has grown in more than a century, accord-

Table 4.1: The official dictionary [Akademio de Esperanto, elaboration]

year	*morphemes*
1905	2,768
1909	3,574
1919	4,161
1921	4,364
1929	4,484
1934	4,492
1935	4,513
1958	4,544
1974	4,744
2007	4,953

ing to the language planning publications made by the *Lingva Komitato* ('Language Committee') and the *Akademio de Esperanto* ('Academy of Esperanto'), an institution similar to the Italian *Accademia della Crusca* or the Spanish *Real Academia Española*. It is noteworthy that these decisions were mostly a posteriori ratification of lexemes already in use – namely, *Oficiala Aldono*, 'official add' – a very light form of corpus planning indeed. There were only three exceptions: in 1923, when the root *analog-* substituted *analogi-*; in 1929, when the prefix *mis-* was added; in 1953, when the suffix *-end-* was added (see chapter 5 for details).[18]

Of course, the dictionary-in-use is broader than the one prescribed by the Academy of Esperanto: a reasonable approximation is given in the recent sociolinguistic presentation by Fiedler [2006] – see Table 4.2.

The number of the Esperanto morphemes is considerably small compared to NLs – for instance, the Oxford English Dictionary lists over 200,000 entries, while the Italian dictionary De Mauro Paravia about 129,000.[19] Nevertheless, Esperanto morphemes are often highly productive. According to Gledhill [1998], the most frequent 6,000 English terms of the American Heritage Dictionary corpus are rendered in Esperanto with about 850 roots.

Last but not least, Esperanto has considerably large linguistic corpora, compared to other IALs. Moreover, they are often available on line for free. In particular, I extracted a set of bitextual examples from *Le Monde Diplomatique*, an international newspaper based in France and regularly translated into Esperanto. In fact, the web editions of *Le Monde Diplomatique* are in 26 different languages at November 2008. Languages vary from Afrikaans to Slovene: see www.monde-diplomatique.fr/int/ for the complete list. The Esperanto web edition has published about 1,000 articles up

Table 4.2: Size of selected Esperanto dictionaries [Fiedler, 2006]

Year	Name of publication	Number of roots
1887	*Unua Libro* [Itô, 1991]	904 roots
1894	*Universala Vortaro*	approx. 2,600 roots
1934	*Plena Vortaro* [Grosjean-Maupin et al., 1956]	6,900 roots + 5,000 compounds
1970	*Plena Ilustrita Vortaro*	approx. 15,250 roots (about 45,000 entries including compounds and derivations)
2002	*La Nova Plena Ilustrita Vortaro* [Duc-Goninaz, 2002]	approx. 17,000 roots (about 47,000 entries)

to now, mostly translated from French. More specifically, the bitextual examples I use are in Italian and English, being the first natural language pair to be used in the translation game as source and target languages.

4.3.3 Esperanto and machine translation

The use of Esperanto in MT is far from being a new idea. In 1933 the Soviet Union Petr Petrovich Troyanskii, "the Babbage of machine translation" (Bar-Hillel), acquired a 'author's certificate', i.e., a patent, for a mechanical translating machine, made of special desk with a typewriter, a photographic camera and a belt as input/output devices. The symbols of logical and etymological parsing were borrowed directly from Esperanto. Afterwards, Esperanto was no longer considered positively in USSR, so Troyanskii dropped its use [Hutchins and Lovtskii, 2000]. Therefore, it can be said that Esperanto was involved in this field before its official beginning, in the 1950s. However, in more recent times other projects were proposed.

The Distributed Language Translation (DLT)

According to the survey by Hutchins and Somers [1992], DLT is the most important MT project using Esperanto:

> In most interlingua-based MT systems, intermediate representations are not genuinely interlingual. Usually the structural representation is language-indipendent, e.g. a predicate-argument structure, but lexical items are not. [...] In DLT, by contrast, the Esperanto interlingua is more like a 'natural language' with its own indipendent structures and lexical items [Hutchins and Somers, 1992, 298].

The DLT interlingua may be well considered the widest, although not complete, formal description of the actual Esperanto syntax we have, and it cannot be disregarded.

DLT started with a seminal study by A.P.M. (Toon) Witkam in 1982, and it was developed for about ten years. A prototype was presented in 1987, while a commercial version was launched in 1993. Both versions had English and French as source and target languages. Esperanto was adopted as the model for the interlingua module until 1988 [Schubert, 1986], [Schubert, 1987], [Maxwell and Schubert, 1989]. The MT engine is written in Prolog.

The syntax is described in terms of a dependency grammar, and a rule system links the different language trees for translation. Schubert [2003] in Ágel and Rau [2003] describes an updated version of this system called 'metataxor' – a term derived from the French *métataxe* used by Tesnière [1959]. According to Hutchins and Somers [1992, chapter 17], the DLT interlingua did not exploited the morphological potentialities of Esperanto, as the MT designers preferred a purely syntactic model.

If semantic disambiguation is needed, as the metataxis had found a valid family of syntactic trees for a given sentence, an apart module called SWESIL (Semantic Word Expert System for the Intermediate Language), would calculate the semantic proximity between the alternatives [Sadler, 1989]. Since 1992, this technique – called 'analogical semantics' – is protected under the U.S. patent n. 5,128,865.

What remains of that project? A recent evaluation of DLT by its seminal leader Witkam points out the following aspects: (a) the choice by Schubert of using of a dependency syntax model for MT was highly innovative at the time; (b) analogical semantics is similar to the example-based approach launched by Nagao [1984] and followed mostly by Japanese researchers [Witkam, 2005].

The Universal Translation Language (UTL)

Sabarís et al. [2001] proposed a project called UTL to be linked with the Universal Network Language (UNL) project, sponsored by the United Nations [Cardeñosa et al., 2005]. UTL used a "constructed human language, based on Esperanto" called Esperanto-UTL which plays the same role as the $\mathcal{L}_e$ in the translation game. Esperanto-UTL shares most linguistic features with standard Esperanto.

Unfortunately, the UTL project seems to be non-active since 2003. Nevertheless, it is a interesting project, as the the authors give some examples of the reduction of structural and lexical ambiguities thanks to the Esperanto very consistent morphology. As a result, the degree of homography is drastically reduced. In fact, homographs may result only from different segmentations of compounds – e.g., *konkludo* may be *konklud-o*, meaning 'conclusion', or *konk-lud-o*, 'play with shells' (see chapter 5 about this topic).

Furthermore, there is almost no allomorphy.

In particular, there are unique final suffixes that mark the main grammar characters, i.e., nouns (*-o* as the subjective singular), verbs (*-i* as infinitive and only five other temporal markers) and their adjuncts (adjectives in *-a* and circumstances in *-e* respectively). For instance, *nordo* means 'the north', *norda* 'northen', *norde* 'at north' and *nordi* 'to be at north'. See Sabarís et al. [2001] for other examples. Esperanto morphology allows for a high degree of productivity from loan translation.

On the other hand, Esperanto-UTL is a different language system, as the authors arbitrarily changed the proposition system in order to enhance the disambiguating properties of their artificial human language, violating the primitive contract already explained in section 4.3.1, and this is the reason behind the fact that no esperantist has shown interest in this project. The standard Esperanto 34-based preposition system is richer and more precise than its source languages [Pennacchietti, 2006a], and hence there is no need to change the delicate linguistic structure of the language. Contra Sabarís et al. [2001], the example 68 – a classic in the literature – shows how the prepositional system of Esperanto is already high precise.

- (68en.) I saw a man with a telescope.
- (68a-eo.) Mi vidis homon kun teleskopo.
- (68b-eo.) Mi vidis homon per teleskopo.
- (68c-eo.) Mi vidis kunteleskopan homon.

Example 68a is the linguistic realisation of the scenario where the seen man had the telescope, while in example 68b the telescope has been used to see the man. As any other (Q)NL, Esperanto allows marked realisations like example 68c, to underline that it is *really* the seen man, the person whom the telescope belongs. In sum, UTL has some good intuition for the general scenario proposed, but it has severe limits in the language planning side of its architecture.

Esperanto in the Visual Interactive Syntax Learning project

In 1996 the Institute of Language and Communication of the University of Southern Denmark has launched a research and development project called Visual Interactive Syntax Learning (VISL) – see the web site visl.sdu.dk. The theoretical base is Karlsson et al. [1995]'s Constraint Grammar parser, a linguistic framework for annotating corpora of different NLs with a general tagset.

In 2003 an on-line corpus for Esperanto was created, annotating the only monolingual corpus available at the time, the *Tekstaro de Esperanto*,

following the Text Encoding Initiative (TEI) standard but without grammatical annotation [Tonkin et al., 2005] – now available at tekstaro.com.

The resulting corpus is about 18.5 million words, with different sources, from the Bible to Wikipedia [Bick, 2007] in [Davies et al., 2007]. It is available for free in the multilingual search engine CorpusEye – see the web site corp.hum.sdu.dk.

With the collaboration of GrammarSoft ApS (Denmark) and Kaldera Språkteknologi AS (Norway) – which offer professional MT and CAT services for Scandinavian NLs – a MT system Danish-Esperanto was released too – available at gramtrans.com.

The purpose of a tagger is to analyse sentences and to add some linguistic information to their constituents, typically words, more rarely morphemes: Warin [2004] compares the Constraint Grammar tagger for Esperanto, which is rule-based, and the gold standard stochastic tagger, i.e., the Hidden Markov Model (HMM) tagger. The rule-based tagger obtained slightly better results.

Other projects and experiences

As far as I know, there are two other Esperanto part-of-speech taggers until now: Minnaja and Paccagnella [2000] and Toral et al. [2005]. The latter in particular is compliant with the EAGLES standard; its tagset is considerably small (86 tags) compared to NLs, e.g., 114 for English, 274 for Italian.

The only open-source project of MT that includes Esperanto as one of the languages is Apertium. Its engine performs a shallow transfer, i.e., the syntax is not analysed deeply but the lexicon is reach enough to permit the right substitutions. This approach proved to be fit for closely related NLs, such as Spanish, Catalan, Occitan and Portuguese, by the Transducens group at the Universitat d'Alacant, and was recently extended to not so related (Q)NLs, such as English, Esperanto, Basque or Welsh. For more information see the official web site apertium.org. [20]

There are two more MT projects using Esperanto. The first one is called Traduku, and it was launched in 2001 by a single programmer. Now, it is publicly testable at traduku.net, but it seems to be not professional. The other one is called Unikomo, and it was launched in 2006 by Bernard Stollman. It follows the transfer model and Esperanto is at its core, while NLs are currently implemented around it. It is too early to write an evaluation.[21]

4.4 Critique of the translation game

As in the classic papers by Turing [1950] and Searle [1980], I will briefly discuss some possible contrary views to the approach presented insofar, along with its counterobjections.

4.4.1 The chinese room argument

> "Whatever linguistic model you put into the machine, we cannot consider it really cognition, as the meaning of the linguistic model is only in the brain of Charles, the system designer".

This argument is based on Searle [1980]. As stated in the introduction, I think that MT is the best testbed of our linguistic models, if they are expressed formally. Even if the linguistic model is only in Charles' brain, the model is made explicit by the computer program. Furthermore, the model will be refined during implementation thanks to the tests given by the computer program itself.

4.4.2 The engineer's reaction

> "Machine translation is not a testbed of any linguistic theory or anything else. What we need is something practical, i.e., commercially valuable as useful in some domains where we want fast translation of large amount of data."

This argument is seldom said openly but nevertheless a lot of people believe in it. I do not agree at all. Having no linguistic theory is a linguistic choice indeed. After all, even some recent developments in statistical MT show the need, or at least the opportunity, to improve the quality of translations through some kind of syntax. Computational brute force and statistics are not enough for achieving good results in MT.

4.4.3 The desperantist's argument

> "Esperanto is not fit for the purpose you mean, as it is an artificial language. You choose Esperanto as the $\mathcal{L}_e$ for propaganda. Why not other international auxiliary languages such as Ido or Interlingua? Why not a logic artificial language as Lojban? It is not better than Klingon!"

People have a bias against Esperanto which is essentially a prejudice, and this is the most clever form I have found to express this common argument. This prejudice is due to a misunderstanding about the term 'natural' in languages: from a theoretical linguistic point of view, Lyons [2006] clarifies four types of naturalness about languages, and hence its antonyms non-naturalness and unnaturalness, until achieving the concept of QNL. However, I have chosen Esperanto for a very practical reason: it is a carefully planned language and it has a considerable amount of corpora available in digital form, compared to competitors such as Ido [Gobbo, 2005a]. Furthermore, unlike any other competitor, it is quite easy to approach the language from most natural languages of the world and a lot of personal experiences agree that everyone may achieve a consistent level of compence

quite quickly. Vice versa, a so-called 'logic' language as Lojban is difficult for Alice to be mastered, at least for the lexicon – which is opaque for everyone.

4.4.4 The typologist's objection

> "Esperanto is fundamentally a European-based language; your scenario may work with English, French or Spanish, but not with non-European languages, such as Chinese, Arabic, or Tamil."

This argument is very serious, and I have only a partial answer. The Esperanto lexicon is essentially pan-European, but its morphology surprisingly resembles non-European language systems as for instance Hungarian or Turkish [Gledhill, 2001]. The choice of English and Italian as the generated natural languages is a real limit, but this is only due to my linguistic repertoire – a pragmatic reason. If I have good results, I try to generalise the adgram model by introducing non-European languages.

4.4.5 The human interface argument

> "Your assumption is too strong. You force Alice not to use her mother tongue, i.e., Spanish, and you ask her to learn Charles' system too. Furthermore, as your approach implies a strong supervision, I think that it will be easier, faster and cheapier to translate source and target language by professionals instead of using your system."

This pragmatic argument moves from economics. It is not completely true that Alice can not use Spanish: she can choose alternative translations of each syntagm as proposed by the system after reading. As claimed in the very description of the scenario, the comparison between the $\mathcal{L}_e$ and the source language should compensate Alice's additional work. First af all, note that the scenario implies a strong supervision only in the analysis phase, and only a monolingual parser is needed, as natural language pairs are generated. Furthermore, the more the TMs are stored, the more precise and wider the system will become. Last but not least, I plan to release the system as an open source project in order to let the worldwide Esperanto community members – always at least bilingual – so to improve the linguistic coverage through new consistent examples and natural language pairs, if a consistent output is obtained.

Notes

[1] For a survey over the theoretical paradigms, Hutchins and Somers [1992] is still valid.

[2] The language policy of the International Academy of Sciences (AIS) of San Marino is based on this principle, and proved to be effective, according to their members [Frank and Fössmeier, 2000].

[3] This is an inheritance of Ceccato's Correlational Grammar: "the operations that consitute this human universal [of mental functioning] are non-linguistic and their products form the substrate of 'meaning' that is designated by different linguistic means" [Glaserfeld, 2001].

[4] The English version of the Common European Framework of Reference for Languages is published by Cambridge University Press. However, it is also available on line. Admittedly, it is hard to imagine a bilingual speaker of Spanish and Tamil; however, this is a thought experiment: Spanish and Tamil were chosen only for the first letter of their respective names in English.

[5] For a long time, the best references about this topic were two volumes: Couturat and Leau [1907] and Couturat and Leau [1903]. In particular, Couturat was influenced by the Leibniz' researches, which he rediscovered in a study trip at Hannover in 1901 – see Couturat [1903] and Couturat [1901].

[6] The term 'planned language' was invented in German as *Plansprache* by Eugen Wüster [Blanke, 1998]. For a full classification of the evolution steps of a IAL, see Blanke [1985], a still valid reference. Recent advancements in the field are published in Blanke [2006].

I prefer the term 'planned language' as a hyperonym of 'IAL' instead of 'interlanguage' or 'interlingua' as this last term has a lot of different meanings: (a) a language variety of L2 learners (i.e., with influences form L1 or overregularisation traits; (b) a formal language in MT systems; (c) 'Interlingua' is the name given by Giuseppe Peano to his IAL *latino sine flexione* in 1909 [Luciano and Roero, 2008]; (d) the name also refers to the IAL of the International Auxiliary Language Association (IALA), as published under the direction of Alexander Gode; (e) a synonym of IAL, as intended by Jespersen in his talk of 1931.

[7] For a survey, see Eco [1993]; for Solresol, see in particular Cherpillod [2008].

[8] The quotation is taken from the cahiers collected by Robert Godel that formed Ferdinand de Saussure's *Cours* in Gobbo [2008a, 27].

[9] Zamenhof as a Jewish philosopher was investigated in particular by Gishron [1986] Maimon [1978] and Holzhaus [1969]. Also see the complete works by Zamenhof edited by Itô Kanzi, alias Ludovikito – see in particular Itô [1982].

[10] There are two key works by Zamenhof to understand his intellectual background: the call to the Jewish intelligentsia and his grammar of Yiddish (written in a Latin alphabet). They were written in Russian; the first one remains a manuscript, the second was published in the Yiddish journal *Lebn un Visnšaft*. Today the originals are in the Jewish National and University library of Jerusalem. They were translated in Esperanto by J. Kohen-Cedek and revised by Adolf Holzhaus, who republished both texts as a single volume in Holzhaus [1982].

[11] The first Esperanto books were recently reprinted as camera-readies in Itô [1991].

[12] For Volapük, see the life and work of Reinhard Haupenthal – see at least Vallon [2006].

[13] For his language policy, see in particular the essays in the collection of Zamehnof's original writings edited by Dietterle [1983].

[14] For a survey of the history of classic IALs, i.e., before the computational turn, see the still valid Large [1985]. The effects of internet to the life of planned languages was investigated in Gobbo [2005b]. Finally, for a comparison between the sociolinguistics of Esperanto, Ido and Interlingua, see in particular Gobbo [2005a] and Gobbo [2008a].

[15] The reference of the history of the persecutions of Esperanto is Lins [1988], available in different NLs. A good survey of the history of Esperanto is Itô [1998], while the only sociological analysis in English of the Esperanto movement is Forster [1982].

[16] An insightful analysis of the whole phenomenon is Lindstedt [2006]. A more technical article about some phenomena of overregularisation is Corsetti et al. [2004].

[17] See for example the multilingual thesaurus *Komputeko*, http://komputeko.net.

[18] These elaboration is based on the statistics published by the Akademio de Esperanto and elaborated furtherly by Kück [2008].

[19] The Frequently Asked Questions of Oxford University Press states:

> The Second Edition of the *Oxford English Dictionary* contains full entries for 171,476 words in current use, and 47,156 obsolete words. To this may be added around 9,500 derivative words included as subentries. Over half of these words are nouns, about a quarter adjectives, and about a seventh verbs; the rest is made up of interjections, conjunctions, prepositions, suffixes, etc. These figures take no account of entries with senses for different parts of speech (such as noun and adjective).

Source: http://www.askoxford.com/asktheexperts/faq/. Retrieved 19 November 2008. In comparison, the Italian dictionary De Mauro Paravia contains 129,432 entries, while de potential derivative forms – in particulary of verbs and nouns – are 734,692 (Pearson Paravia Bruno Mondadori spa, Fabrizio Cicoira, personal communication, 19 November 2008). However, it is highly debatable how to count words, exactly for the existence of morphology: *admitt-able* is distinct from *admitt*? And what about *water-fall*? Moreover, a stative use is to be counted separately from a adjective or a circumstantial use, e.g., *wrong*? There is no general agreement among specialists about this topic, and in any case this problem is out of the scope of this dissertation: let us consider these numbers simply as rough estimates.

[20] I am collaborating informally for Esperanto and Italian – see the official wiki of the project.

[21] I am collaborating informally with the Unikomo project concerning the Italian language.

Chapter 5

The adpositional grammar of Esperanto

The core of the MT scenario proposed in the previous chapter is the formal description of the Esperanto grammar. In this chapter, the adpositional grammar of Esperanto is provided, for two reasons. The first one is to give a concrete instance of the general model of adtrees and dictionary described respectively in chapters 2 and 3. The second reason is to give an instance of the translation game in vitro: for this reason an apart attention is given to parallel examples in English and Italian as the source and target NLs of the very system, although other parallels in other NLs are used when needed.

Being Esperanto built over NL various material, it is important to make a distinction clear between **Esperanto source languages** and other NLs. Zamenhof's linguistic repertoire was: Yiddish and Belarussian (his mother tongues), Polish, German, Lithuanian (corresponding to the nationalities of his surrounds), French (the international language of his time), English (self- studied), Greek and Latin (high studies) and Hebrew and Aramaic (Talmudic studies).[1] These 11 NLs form the set of Esperanto source languages. The relation between Esperanto and its source NLs is similar to the relation between Latin and Italian, or Old Saxon and contemporary English: they are the language substrata of Esperanto. This means that Esperanto borrows their structures – i.e., phonology and morphosyntax – in an original combination. Therefore, it is different to say that Esperanto has a Saxon genitive case, although fossil, as in English, from saying that its morphology is mainly agglunative as in Turkish or Japanese. The parallel with English is stronger, as English is a source language, while the parallel with Turkish or Japanese can be interesting, but in any case it is accidental, as Turkish and Japanese are not source languages, i.e., these similarities have been unplanned by Zamenhof. Nonetheless, if there would be a consistent percentage of Japanese or Turkish Esperantophones in the worldwide community, this could lead to a significant influence by those

NLs. However, in the history of Esperanto, by now the most influencing national group was the French, especially between the two world wars, and this had have a notable influence, especially in the borrowing of new words.

Borrowing is a crucial aspect in Esperanto. In fact, the meaning of lexemes, even if borrowed from source languages, can be different from the original, as Esperanto is an autonomous linguistic system. Now, a specification is needed. I call **calque** a loanword where neither phonetics nor morphosyntax is respected in the target language, but only semantics. In contrast, an **importation** is a loanword where graphemes are respected into the target language, while semantics can be respected or not.[2] For example, the French word *affaire* is borrowed as an English calque as *affair*, while the French importation *affaire* has a restricted meaning, namely a *love* affair. Esperanto generally preferes calques to importations; however, Esperanto calques can have a different meaning compared to the source languages, exactly as Old English words can have different meanings compared to contemporary English. For example, the Esperanto word *afero* ('affair') has an additional meaning of 'the affair(s) concerning Esperanto, i.e., the straight engagement in developing the language by using it and make it be used by others'.

Esperanto morphosyntax is highly flexible, as already shown in example 68 (see section 4.3.3). Its unique features are suitable for the L_e role in the translation game. For example, the construal of the concept 'to have a shower' is different between English and German or Italian, and Alice, the MT engine user, would express it accordingly.

- (69en.) Alice had a shower.
- (69de.) Alice duscht.
- (69it.) Alice fece una doccia.
- (69a-eo.) Alico havis duŝon. (English-like)
- (69b-eo.) Alico faris duŝon. (Italian-like)
- (69c-eo.) Alico duŝis. (German-like)

The series 69a-b-c is perfectly grammatical in Esperanto, while the German-like Italian construal **Alice si docciò* or the Italian-like English construal **Alice did a shower* are not.[3] This is a great help to the MT engine user. Of course, the price of this flexibility is a considerable complexity degree in collocation, which is by no means simple or easy, as shown in the seminal work by Jansen [2007].

In the following I present the essential adpositional grammar of contemporary, standard Esperanto, with reference to the classic text of the origins for control. In other words, first I will base myself on my linguistic

ability, being an **Esperantophone**, i.e, an active speaker of Esperanto, and an **Esperantist**, i.e., a member of the worldwide community of supporters. Second, I use recent monolingual and bilingual dictionaries and grammars in various NLs (mainly English, Italian, German) for reference.[4] Third, I use some studies about the language use of Zamenhof for double-checking, so that evolutional traits can emerge for contrast.[5] Caveat: what follows is a description of Esperanto for machine translation purposes, not a grammar for Esperanto learners. This means that every concept is written for the purpose of implementation in the formal model described linguistically in the first part of this dissertation, and formally and computationally since here. In particular, while the adspace is covered totally, only some affixes are described in their moduli and use. Readers interested in learning Esperanto should turn their attention to the many handbooks and e-learning courses available in many NLs.

5.1 The morphosyntax of Esperanto

In adgrams, the dictionary is made by morphemes, and their rules of combinability and collocation in the syntagmatic axis form morphosyntax. Therefore, the first step to describe the grammar of a (Q)NL is to classify the morphemes under the cross-linguistical classes already presented in section 3.4: adpositions from one side and lexemes from the other side, along with their derivations, i.e., transferers and tranferees. I recall that transferers are a subset of adpositions, while transferees are a subset of lexemes. The second and last step is to give language-dependent subclasses according to the semantic maps designed by the (Q)NL, so to build the dictionary suitable for computational traitment. Table 5.1 shows the general morphological classes, while Table 5.2 shows transferers and transferees. These classes are cross-linguistically valid. Each morpheme belongs to one and one only class.[6]

Esperanto has a rich inventory of about 40 affixes, mostly prefixes and suffixes, which are of wide use. These inventory was set since the beginning by Zamenhof, and afterwards they were approved and monitored by the Academy of Esperanto, with very few changes – see chapter 4 for details. Wennergren [2005, §39] correctly reports and analyses in details unofficial affixes too, which are used mainly for specific domain needs, such as zoology. For the purposes of this dissertation, the affixes and their conventions presented hereafter are only the official ones. Furthermore, some prepositions and a pronoun can be used as affixes: they will be analysed too apartly, because their syntactic value is different.

Affixes always add moduli to the lexeme to which they are attached: the affix is the dependent, while the lexeme is the governor (see chapter 2). The dependency between affixes and lexemes is completely grammaticalised,

Table 5.1: General morphological classes

Adpositions		**Lexemes**	
basic level		*selective*	
Plus	$\oplus$	verbs	(I)
Minus	$\ominus$	adverbs	(E)
Slash	$\oslash$	nouns	(O)
Times	$\otimes$	adjuncts	(A)
phrasal level		*collocational*	
Plus	$\oplus$	verbal l.	(I> x)
Minus	$\ominus$	circumstances	(E> x)
Slash	$\oslash$	stative l.	(O> x)
Times	$\otimes$	adjunctive l.	(A> x)

Table 5.2: The special lexemes: tranferers and transferees

Transferers		**Transferees**	
verbifying	(x >I)	verbal	(I< y)
circumstantial	(x >E)	circumstantial	(E< y)
stative	(x >O)	stative	(O< y)
adjunctive	(x >A)	adjunctive	(A< y)

Table 5.3: The Esperanto alphabet

a, A	b, B	c, C	ĉ, Ĉ	d, D	e, E	f, F
g, G	ĝ, Ĝ	h, H	ĥ, Ĥ	i, I	j, J	ĵ, Ĵ
k, K	l, L	m, M	n, N	o, O	p, P	r, R
s, S	ŝ, Ŝ	t, T	u, U	ŭ, Ŭ	v, V	z, Z

hence their adtype is always Slash (⊘).

Some affixes can be applied only if the lexeme has a precise grammar character, so they will be presented according to the grammar character of the lexeme they are applied to. In Esperanto, the only transferers are the fundamental moduli, two special valence suffixes and the compound rules (see below). Hence, *affixes are transferees and therefore part of the lexicon*. In fact, quite often Esperanto affixes are used as lexemes, i.e., without a governor: *-ebl-e*, 'perhaps', *-ar-o*, 'a set', etc. This is not a problem in adgrams: simply that affix will be the lexical head of its adtree, i.e., it will put as the rightmost participant.[7] In Esperanto, it is possible to attach more than one affixes to a given lexeme, like other agglutinative NLs such as Turkish, Hungarian or Japanese: for this reason, they should be parsed accordingly to their value (see section 6.3 in the next chapter).

In the following, the adpositional space of Esperanto will be presented at first, then the lexicon will be covered. The order of presentation of lexemes and affixes will follow the grammaticalisation degree already seen in section 3.3: I → E → O → A.

5.2 The Esperanto alphabet

Before to proceed, it is necessary to introduce the Esperanto alphabet, which is derived from Latin – see Table 5.3 (for pronunciation, Wennergren [2005, §2]). As in the case of many NLs, Esperanto has its own unique combinations letter plus diacritic, which act as identity flag, such as the Spanish {ñ} or the German {ß}. The four letters {q}, {w}, {x}, {y} are used only for transcribe imported foreign names, e.g., 'Yvonne'.

Each phoneme is written through one only grapheme: in other words, Esperanto follows a graphemic writing system, unlike for example English which follows an etymological one.[8] The Esperanto alphabet is easily encodable through the Unicode Standard.

5.3 The adpositional space of Esperanto

Adpositions in Esperanto are very clear for two reasons. The first reason is the presence of shallow signatures of the fundamental moduli, and the second one is the almost absence of allomorphs. This fact does imply that the adpositional space of Esperanto is neither small nor trivial: in the following, I present the conditio sine qua non so to treat Esperanto adgrammatically in parsing, adtree generation and adtree transformation (i.e., Tesnière's metataxis) for machine translation.

5.3.1 The fundamental moduli

Esperanto has the unique property of having explicit signatures of the fundamental moduli (see Table 3.4). This means that the number of transferers is very, very small compared with NLs, as there is no need to resolve complex transfer chains. Lucien Tesnière explicity acknowledges his debt to the prototypical Esperanto signatures of grammar characters, which are a crucial part of his grammar system, in a note that almost nobody was ever aware of, even in the community of dependency grammar specialists.

> This particularity is a useful mnemonic tool, as Esperanto is more or less an intermediate of diverse European languages [Tesnière, 1959, my translation].[9]

Unlike other NLs, the Esperanto set of final affixes explicits *the final grammar character of the current group, regardless of the original grammar character of the proper lexeme of that group*. Let me explain through example 62, already presented in English in Figure 3.8 (section 3.4.3). The

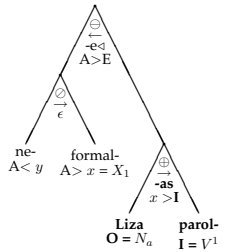


Figure 5.1: The adtree of *Liza parolas neformale* (62).

main difference between English and Esperanto is in the lexeme *formal-*, which is a selective adjective in Esperanto, while *formal* in English is derived from the stative lexeme *form*, while *-al* is an adjectival transferer.[10] In order to be transferred to a circumstantial, the stative lexeme *form* should

Table 5.4: Signatures of the fundamental moduli

$V_{\text{ERBIFICATION}}$	$C_{\text{IRCUMSTANTIATION}}$	$A_{\text{DJUNCTIVATION}}$	$S_{\text{TATIVATION}}$
I	E	A	O
-as	-e	-a	-o
-is	-en	-aj	-oj
-os		-an	-on
-u		-ajn	-ojn
			-i

be transferred in an adjective before: in fact, *form-al-ly* (O>A>E) is right, while **form-ly* (O>E) is ungrammatical. I call transfers such as (O>A> y) in the English example 62 **inline transfers**, while transfers such as (x >A>E) **final transfers**. A **transfer chain** is a chain of a series of inline transfers (which can be empty) and a final transfer (always one, mandatory). The English language is full of such tranfer chains O>A>E: *mathematic-s : mathematic-al : mathematic-al-ly*; *region : region-al : region-al-ly*; *nation : nation-al : nation-al-ly*. Of course, in many cases the transfer chain is not evident, often because of the rich etymological table of English: *hell : hell-ish/infern-al : hell-ish-ly/infern-al-ly*; *bas-e : bas-ic : bas-ic-al-ly*. In contrast, Esperanto performs transfers O>E directly to the final grammar character, without passing through an adjectival transferer: *matematik-o : matematik-a : matematik-e*; *region-o : region-a : region-e*; *naci-o : naci-a : naci-e*; *infern-o : infern-a : infern-e*; *baz-o : baz-a ; baz-e*.

In Esperanto, inline transfers do exist, but they can be almost not computed at all, as the final transfer is always explicit, thanks to the signature of the fundamental moduli.[11] Table 5.4 shows the Esperanto signatures of the fundamental moduli, i.e., *all Esperanto final transferers*. They are compulsory for every lexeme, even if redundant. For example, the lexeme *elefant-* is clearly stative, nonetheless it should be always marked with a signature of stativation to be actually used, e.g., *elefant-o*, 'an elephant' (see section 5.5.1 below for details). This highly simplifies the implementation of parsing.

5.3.2 The quasi-symmetry of adjunctivation and stativation

Some of the signatures presented in Table 5.4 are simple, i.e., composed by one morpheme, while others are complex, i.e., composed by more than one morpheme. In particular, adjuncts and stative lexemes are almost symmetrical. Table 5.5 shows the basically agglutinative character of Esperanto, as each single adposition brings one and only modulus: stativation is marked by the final *-o*, adjunctivation by *-a*, plurality by *-j* (section 5.3.4) while fi-

Table 5.5: Symmetry of the proper adjunctive and stative signatures

O	A	NUMBER	CASE
-o	-a	singular	nominative
-o-j	-a-j	plural	nominative
-o-n	-a-n	singular	accusative
-o-j-n	-a-j-n	plural	accusative

nally the marked case, i.e., accusative by *-n* (section 5.3.5). The symmetry in adjunctivation and stativation is due to agreement: the symmetry is broken by the infinitive (see section 5.3.3 below).

The signatures of the fundamental moduli can be redundant of the grammar character of the lexeme, e.g., *elefant-o* ('elephant'), *blu-a* ('blue'), *kur-as* ('run'), or act as transferers, e.g., *elefant-a* ('elephant-ine'), *blu-o* ('blue-ness), *kur-o* ('a run'). This redudancy will be treated later in details.

5.3.3 The basic form of a verb is stative

Esperanto borrows the infinitive from Latin deponent verbs, e.g. *loqu-i*, 'to speak, to talk'. As in English or Italian, the infinitive is used to indicate the basic form of the verb, and that is why Tesnière have chosen the letter I for verbification. However, from a syntactically point of view, an infinitive is *not* a verb: i.e., its valence(s) are not active, and it acts as a vehicle of actants. Example 70 shows different construals for the same concept.

- (70a-eo.) Mi ŝatas danci.
- (70a-en.) I like to dance.
- (70b-eo.) Al mi plaĉas danci.
- (70b-it.) A me piace danzare.
- (70c-eo.) Mi plezuras per danci.
- (70c-it.) Io mi diverto a danzare.

The English construal is in example 70a, where the act of dancing is the syntactic object O, while the Italian construals are exemplified by 70b, where the act of dancing is the syntactic subject S. Finally, in example 70c the Italian construal represents the act of dancing as an indirect object, i.e., a stative group introduced by a preposition. Esperanto respects the mental habits of English and Italian speakers, only with small syntactic adjustments (Figure 5.2. In any case, the infinitive cannot have a verbal (i.e., V^{0-3}) value. Table

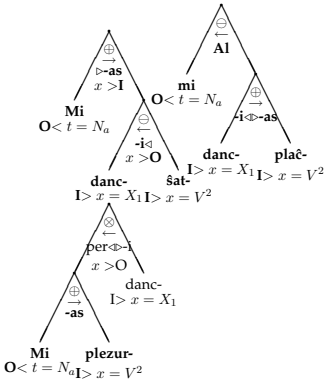


Figure 5.2: The adtrees of *Mi ŝatas danci* and the like (70).

Table 5.6: Number and gender in the infinitive signature

infinitive	NUMBER	CASE
-i	any	any

5.6 shows that the infinitive does mark explicitly neither the number (singular vs. plural) nor the case (nominative vs. accusative). In contrast these moduli are marked explicitely by stative lexemes and adjuncts.

5.3.4 Singular and plural

Esperanto has only one signature for the modulus +[PLURAL], i.e., *-j*, which is attached immediately after stative lexemes and to adjuncts. Verbal lexemes or circumstantials never have the plural modulus.

- (71a-eo). Junaj knaboj kuras rapide.
- (71a-en). Young boys run quickly.
- (71a-it). I ragazzi giovani corrono velocemente.
- (71b-eo). Juna knabo kuras rapide.
- (71b-en). A young boy is running quickly.
- (71b-en). Un ragazzo giovane sta correndo velocemente.
- (71c-eo). Ambaŭ knaboj kuras rapide.
- (71c-en). Both boys are running quickly.
- (71c-en). Entrambi i ragazzi stanno correndo velocemente.

In example 71a, the stative group *knaboj* 'some boys' and the adjunctive one *junaj* 'young' are both marked with the final *-j*, while in 71b – Figure 5.3 –

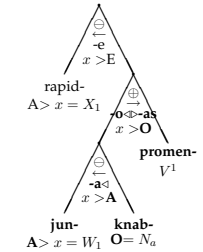


Figure 5.3: The adtree of *Juna knabo promenas rapide* (71b).

they are not.[12]

Furthermore, there is a fossil dual form, *ambaŭ*, 'both', as shown in example 70c. Finally, there is an explicit marker of collective nouns, i.e., *ar-o*: *hom-ar-o*, 'human-kind', *gazet-ar-o*, 'the newspapers', *arb-ar-o*, 'a forest, a wood'. In this case, syntactically these names are singular even the concept referred are plural.

5.3.5 The accusative

The basic word order of Esperanto is clearly SVO.

- (71d-eo). Junaj knaboj sekvas rapidan hundon.
- (71d-en). Some young boys are following a quick dog.
- (71d-it). Dei ragazzi giovani seguono un cane veloce.
- (71e-eo). Juna knabo sekvas rapidajn hundojn.
- (71e-en). A young boy is following some quick dog.
- (71e-it). Un ragazzo giovane segue un cane veloce.

The presence of an explicit marker of the syntactic object O, i.e., the morpheme *-n* let a certain degree of freedom:

> [Esperanto] show[s four] basic sequence[s] of sentence constituents which can be expressed syntactically. The first of these concerns the positions of the nominal subject S and the nominal direct object O in relation to the verb V in neutral main clauses. The basic order thus defined is SVO. [...] Under the influence of the pragmatic functions of focus and topic, S, V and O all exhibit such mobility that no less than 10 per cent of all their combined occurrences deviate from the basic order SVO. The peripheral positions in the sentence appear more suitable for placing constituents in foucs. Invertion of constituents, aimed at focussing on one or the other, proves to be commoner than the use of the cleft sentence. There is a clear tendency to follow the so-called LR Principle, expressed in the form of the Theme-Rheme (ThRh) or Topic-Comment (TC) sequence. Inversion of constituents for the purpose of bringing the topic forward is very common. [Jansen, 2007, 273–274, references omitted].

The accusative is always a Slash (⌀) adposition, unless if used in combination with prepositions: in those cases the adtype is decided according to the preposition (see section 5.3.8 below).

The morpheme *-n* is directly borrowed from the German (and Yiddish) accusative, and it has very similar functions [Willkommen, 2007, 27].[13] In stative and adjunctive groups, it is always applied immediately after the plural, if present, or after the stative and adjunctive signature: *jun-a*, *jun-a-j*, *hund-o-n*, *hund-o-j-n*. It works analogously with stative and adjunctive correlatives (see respectively sections 5.3.9 and 5.8.2). The fact of having an explicit marker for the accusative case permits to be clear in sentences where NLs as Italian are ambiguous. Let me recall and expand example 61.

- (61ab-it.) Paolo ha trovato il telefono rotto.
- (61a-eo.) Paŭlo trovis la telefonon rompitan.

- (61a′-eo.) Paŭlo trovis la rompitan telefonon.
- (61a-en.) Paul has found the broken telephone.
- (61b-eo.) Paŭlo trovis la telefonon rompita.
- (61b-en.) Paul has found the telephone out of order.

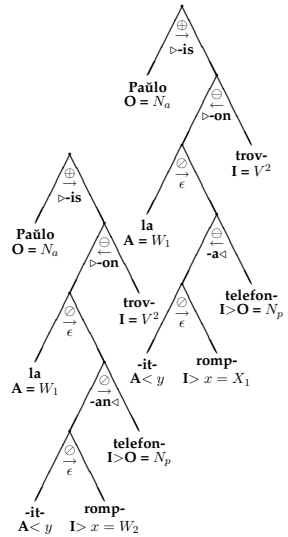


Figure 5.4: Predicative and attributive adjuncts in Esperanto (61ab).

Figure 5.4 shows the Esperanto examples 61 (61a and 61a′ on the left, 61b on the right). In this example Esperanto is very similar to Italian, but without ambiguity (see Figure 3.7 in section 3.3.4). Of course, this does not mean that Esperanto is a 'perfect language', i.e., it has no ambiguities, it never gives a family of adtrees as the output, etc. It does only mean that Esperanto does not lack a certain degree of accuracy in construals.

The basic modulus of the morpheme *-n* is to sign the syntactic object O, which I sign as +[N] for conciseness. However, it can also bring two additional moduli: +[MOVE] and +[MEASURE].

- (72a-eo). Morgaŭ mi veturos Parizon.
- (72b-eo). Morgaŭ mi veturos al Parizo.
- (72ab-it). Domani andrò a Parigi.
- (72ab-en). Tomorrow I will ride to Paris.

The modulus +[MOVE] can be activated only if the stative group has the modulus +[PLACE] active, as stated by example 73 [Wennergren, 2005, from §12.2.5].

- (73a-eo). *Mi iras kuraciston.
- (73b-eo). Mi iras al kuracisto.
- (73ab-it). Vado dal dottore.
- (73ab-en). I go to the doctor.

The other use of accusative is with the modulus +[MEASURE], which inherits the modulus +[DEFINITE]. This modulus can be used for its two prototypically subclasses +[TIME] and +[SPACE] or any other.

- (74-eo). Pluvis unu tagon.
- (74-it). Piovve un giorno intero.
- (74-en). It had rained for one day.
- (75-eo). La dorsosako pezas kvin kilojn.
- (75-it). Lo zaino pesa cinque chili.
- (75-en). The backpack weighs five kilos.
- (76-eo). La libro kostas dek eŭrojn.
- (76-it). Il libro costa dieci euro.
- (76-en). The book costs ten euros.

Last not least, there is an 'accusative of effect', sometimes called 'Greek-like accusative' in grammars of Latin. It is used very rarely, and only in literature. The stative group *kap-o-n* of example 77 is an example of its use. [Wennergren, 2005, from §12.2.6].

- (77-eo). Ĉu vi permesos al mi pendigi ĉi tiun kanajlon sub la ĉielo la kapon malsupre?
- (77-en). Mi permetti di appendere questa canaglia sotto il cielo, la testa di sotto?
- (77-en). Do you let me hang this blackguard under the sky, his head below?

Table 5.7: Circumstantial finals in Esperanto

circumstantial	E	TOWARDS
-e	⊤	⊥
-en	⊤	⊤

5.3.6 The circumstantials

Unlike English or Italian, in Esperanto the signature of circumstantials are trivial. The final *-en* is a circumstantial accusative (see below). There is no marking of the number, i.e., **-ejn* is ungrammatical. There are some adverbs in Esperanto, i.e., selective lexemes with a circumstantial value (see section 5.5.3). Furthermore, there are some correlatives with a circumstantial value (see section 5.3.9).

The circumstantial accusative

The combination circumstantial+accusative, i.e., *-e-n* activates the moduli +[MOVE,TOWARDS], the letter being a subclass of the first one.

- (72c-eo). Morgaŭ mi veturos Parizen.
- (72c-en). Tomorrow I will ride towards Paris.

In example 71c, the Agent rides towards Paris without necessarily achieving it, while in 71a the Agent certainly arrived [Willkommen, 2007, 66].

5.3.7 Verbification

Esperanto has five verbal transferers. Verbal finals in Esperanto are trivial. Every verb follows the same paradigm, even 'to be' and 'to have', as Esperanto is a carefully planned language.[14] Verbal finals change neither for the person nor for the number. Verbal finals are in position I_0 in parsing. Table 5.8 shows the correspondence between each final and their moduli. Some esperantologists argue that the verbal finals bring only tense, not aspect. As shown by Wennergren [2005, 380], this is not true: the Zamenhofian sentence *En la vintro oni hejtas la fornon* ('In winter it is used to warm the oven up') shows clearly the modulus [HABIT]. Of course, if the modulus [HABIT] is active the modulus [PROGRESSIVE] will be not. Analogously, in non-marked sentences the form in *-os* signs the belief of the speaker that the construal will really happen in the world model of reference. Interestingly, the *-us* final is antonym both of [PRESENT] (*-as*) and of [BELIEF] (*-os*). Mostly it is used with the junctive *se* ('if').

Table 5.8: Verbal finals in Esperanto

Final	Tense	Aspect
-as	+[PRESENT]	+[PROGRESSIVE]
		∪+[HABIT]
-is	+[PAST]	+[PERFECT]
-os	+[FUTURE]	+[BELIEF]
-us	−[PRESENT]	−[BELIEF]
-u		+[IMPERATIVE]
		∪[+[DESIRE]]
		∪[+[WISH]]
		∪[+[ORDER]]
		∪[+[GOAL]]

- (78-eo). Se mi estus sana, mi estus feliĉa. [Wennergren, 2005, 383]
- (78-en). If I would be healthy, I would be happy.

It is possible to change tense and aspect through the use of the participles – see section 5.6.5 below. Finally, the form in *-u* is borrowed from English, as the second person (singular and plural) can be omitted.

- (79-eo). Ludoviko, donu al mi panon.
- (79-en). Ludwig, give me bread.
- (79-it). Ludovico, dammi il pane.
- (80-eo). Ni legu la unuan ĉapitron, mi petas.
- (80-en). Please, let's read the first chapter.
- (80-it). Per favore, leggiamo il primo capitolo.

In example 79, the activated modulus is [WISH], while in example 80 the modulus is [DESIRE] [Wennergren, 2005, 382]. Note, that the English form *Let's* for the first person plural is rendered with the pronoun *ni* (see section 5.7 for pronouns). The idiomatic expression *mi petas* (literally, 'I ask'), is used to mild the verbal *-u* final, so it should be treated not as a phrase but as a circumstance of the verb in *-u*, as the English *please* or the Italian *per favore*.

5.3.8 The prepositions

Pennacchietti [2006a] is the most complete study of the Esperanto prepositions in terms of adspaces. A preposition is a particular type of adposition.

It does not participate in transference, hence it is a transferee, but it indicates the adtype and therefore the structure of the adtree. Some prepositions can be applied in a basic or phrasal level, i.e., both as 'proper' prepositions or as junctives. They will be treated here, while section 5.3.10 will be dedicated to 'proper' junctives.

In Esperanto – as in most Indo-European languages – prepositions can be applied only to stative (O) and adjunctive (A) groups. If applied to circumstantial (E) groups, they are grammaticalised as prefixes (see below). Adjunctive groups can be assimilated to stative ones, so only verbal and stative groups will be indicated.

I call the left participant of the prepositional adposition the **governor** while the right participant is the dependent: e.g., *libro de Petro en biblioteko*, 'Peter's book in a library', the governor is *libro* while *Petro* is the dependent; analogously, *libro de Petro belas*, 'Peter's book is nice'. In the following, the grammar characters of governors in prepositional adpositions will be indicated, as their moduli and adtree structure can change. In fact, some prepo-

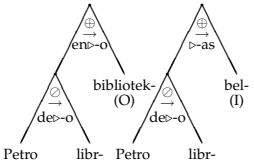


Figure 5.5: Prepositions and their governors (example).

sitions, like *de*, accept both stative governors (O, Figure 5.5, left) and verbal governors (I, Figure 5.5, right). Prepositions that accept only verbal governors can be considered rightly **circumstantial prepositions**, while the ones which accept only stative governors can be called **adjunctive prepositions**.

Referring to Figure 2.4 based on Pennacchietti [2006a], my elaboration of the prepositional space of Esperanto is as follows:

- **Plus**: *en*, *dum*, *tra*, *inter* and *sur*, *super*, *por*, *trans*, *preter*, *antaŭ*, *kontraŭ*, *ĉirkaŭ*;
- **Minus**: *al*, *je*, *ĝis*, *ĉe*;
- **Slash**: *de*, *el*, *da*, *por*, *pro*, *pri*, *sen* and *krom*, *ekster*, *anstataŭ*, *malgraŭ*, *far*;
- **Times**: *kun*, *per*, *laŭ*, *apud* and *aŭ*, *kaj*, *sub*, *post*, *malantaŭ*.

Table 5.9: The Esperanto preposition *de* and its class

Prep.	Governor	Basic modulus	Additional moduli
de	O	+[APART]	+[BELONG]
de	O	+[APART]	+[A]
{de|da}	O,+[MEASURE]	+[APART]	
(ek)de	I	+[APART]	+[TIME,BEGIN]
de (post)	I	+[APART]	+[TIME,END]
de(post)	I	+[APART]	+[TIME,END]
de	I	+[APART]	+[SPACE]
{de|el}	{O|I}	+[APART]	+[SPACE,ORIGIN,FROM]
(dis)de	{O|I}	+[APART]	+[SPACE,MOVE,FROM, AWAY,FAR]
(for) de	{O|I}	+[APART]	+[SPACE,MOVE,FROM, AWAY]
{de|pro}	I	+[APART]	+[CAUSE]
{de|fare de|far}	{O|I}	+[APART]	O= N_a
L_{exeme}-e de	{O|I}	+[APART]	from L_{exeme}
de	{O|I}	+[APART]	

The prepositions listed after 'and' are treated as 'neutral' in Pennacchietti's view for mnemonic reasons, but in the actual version of mine there is no space for neutral subspaces (see chapter 2 for more details).

Before to proceed, it is noteworthy to remember that there are basic-level adpositions which are not prepositions, such as verbal finals – shown in Table 5.6 – are Plus or Minus in case of mirroring, while adjunctive finals – shown in Table 5.5, or stative finals – they can be only Minus or Slash. See again Table 3.9 for the complete picture. In the following subsections, prepositions are presented in contrastive pairs or groups, according to their use, and their appropriate subspace will be indicated, along with appropriate examples, mostly from Wennergren [2005].

Grammaticalisation: *de* and its class

The preposition *de* is the most grammaticalised, even among non-dimensional retroapplication (Slash), and hence its meaning varies at most, as shown in Table 5.9. Prototypically, *de* is Slash and it brings the modulus [APART]. It can depend on an O-group, but it can also be attached to an I-group, with different moduli.

A basic use of *de* is to sign belongingness: *la ĉapelo de patro*, 'father's

cap', *la dentoj de leono*, 'lion's teeth'. Another basic use is to transfer the dependent stative group in an adjective one, i.e., O>A, exactly as the *-a* final (modulus [A]): *virino de meza aĝo* is equivalent to *mezaĝa virino*, 'a middle-aged woman', while *figuro de triobla grandeco* is equivalent to *trioble granda figuro*, 'a three times big figure' (note that *tri-obl-a* has been tranferred to *tri-obl-e*, i.e., an E-group dependent upon an A-group). Nonetheless, this transference is rarely usable: *pomo de oro* is not equivalent to *ora pomo*; the first one is an apple made of gold, the second one a golden apple.

If the governor is a stative group with a modulus of [MEASURE], sometimes it can be substituted with a more narrowly defined preposition, *da*. For example, *la knabo havis la aĝon de nur ses jaroj*, literally '?the boy had the age of only six years', which is hardly acceptable in English but normal in Italian, *il ragazzo aveva l'età di soli sei anni*. In this cases no substitution is possible. However, [MEASURE] is also the explicit, basic modulus of a correlative series – see section 5.3.9: in that case, it can be always substituted by *da*: *kiom de la ŝtofo ili jam pretigis* is equivalent to *kiom da ŝtofo ili jam pretigis*, 'how much fabric they have already prepared'.

The preposition *de* can be attached directly to verbal groups. A common use of *de* is with the additional modulus [TIME]. It can be used to sign the beginning (example 81) or the ending (example 82) of an action – *depost* is Zamenhofian use, but today's Esperanto is *de post*.

1. (81-eo.) Mi promesas (ek)de hodiaŭ ne fumi plu.
2. (81-en.) I promise not to smoke anymore from today.
3. (82-eo.) De(post) la tago, en kiu Varsovio salutis la novan jaron, pasis ses semajnoj.
4. (82-en.) From the day, in which Warsaw had greeted the new year, six weeks passed.

Another common use of *de* is with the additional modulus [SPACE], with different additional moduli.

1. (83-eo.) Mia vojaĝo de Parizo al Londono estis tre laciga.
2. (83-en.) My trip from Paris to London was extremely tiring.
3. (84-eo.) Distingado de bono disde malbono ne ĉiam estas facila afero.
4. (84-en.) To distinguish the good from the bad is not always easy.

In example 83, the prepositions *el*, or *for de*, can also be used. An additional use is with the modulus [CAUSE], which is also a modulus of a correlative series – see section 5.3.9 below.

1. (85-eo.) Mi eksaltis de surprizo.

2. (85-en.) I jumped for surprise.

In example 85 the preposition *pro*, which is highly specific, can also be used. An important use is the signature of Agentivity, both in active (example 86) and in passive (example 87) phrases.

1. (86-eo.) Mi ricevis libron de mia patro.
2. (86-en.) I received a book from my father.
3. (87-eo.) Ŝi estas amata de ĉiuj.
4. (87-en.) Shi is loved by everybody.

The preposition *de* raised a limited but significant list of complex prepositions with very specific meaning during the evolution of Esperanto: *dekstre de*, 'at the right of', *escepte de*, 'except of', *inkluzive de*, 'including', etc. [Wennergren, 2005, 143]. Finally, the most grammaticalised use is completely generic, i.e., without additional moduli:

1. (88-eo.) Saĝa filo lernas de la patro.
2. (88-en.) A wise son learns from his father.
3. (89-eo.) Niaj okuloj estas plenaj de larmoj.
4. (89-en.) Our eyes are full of tears.

The punctuals: *al* vs. *je*

The puntual prepositions *al* and *je* are Minus. The preposition *al* has the precise meaning of showing direction towards a single point as a basic modulus. Table 5.10 shows the current uses of these two prepositions. The default use of this preposition is rather limited, eventually for the influence of its source language substrata: in Esperanto you go *al kongreso*, 'at the congress', *al universitato*, 'at the University', but never *?al ofico*, '*at the office', or *?al hejmo*, even if there is no violation transpassed, theoretically speaking. In contrast, the verbal use of *al* is very, very common, especially with trivalent verbs, where the preposition clearly introduced the Experiencer (N_e), as in examples 90 and 91.

- (90-eo.) Donu al la birdoj.
- (90-en.) Give (it) to the birds.
- (91-eo.) Pardonu al mi.
- (91-en.) (I beg your) pardon.

Table 5.10: The Esperanto punctual prepositions *al* and *je*

Prep.	**Governor**	**Basic modulus**	**Additional moduli**
al	O	+[-WARD,POINT]	
al	O	+[-WARD,POINT]	+[GOAL]
al	O	+[-WARD,POINT]	+[SPACE,MOVE,END]
al	O	+[-WARD,POINT]	+[TIME,MOVE,END]
L_{exeme}-e al	{O|I}	+[APART]	from L_{exeme}
al	I	+[-WARD,POINT]	O= N_e
je	O	+[POINT]	
je	O	+[POINT]	−[DEFINE],+[TIME]
je	O	+[POINT]	−[DEFINE],+[MEASURE]

Other uses of *al* are: to sign the goal (example 92), the end of a movement (example 93), and finally the end of a time period (example 94; *de tempo al tempo*, 'sometimes', is an idiomatic expression).

- (92-eo.) Vi devas nur iri al la fonto.
- (92-en.) You only have to go to the source.
- (93-eo.) Li rigardis al la klara, hela luno.
- (93-en.) He looked at the clear, bright moon.
- (94-eo.) Mi rigardas televidon de tempo al tempo.
- (94-en.) I look at the tv sometimes.

Similarly to the case of *de*, a very generic manner to use *al* is in complex prepositions: *fronte al*, 'in front of', *responde al*, 'answering (to)', *rilate al*, 'in relation with', *spite al*, 'in spite of', *danke al*, 'thanks to'.

Also the case of *je* is of high interest: this preposition was indicated in the first times as a 'jolly', i.e., 'use it whenever you do not know which preposition you should use but you feel you should use one'. With the evolution of Esperanto, its use was naturally restricted to the semantic domain of punctuality. Often the boundaries of a point are flou, i.e., they are not defined, borderless. In examples 95-96, the modulus TIME is active, while in example 95, the modulus MEASURE is active.

- (95-eo.) Ŝi venas je la kvina (horo).
- (95-en.) She comes at five (o'clock).
- (96-eo.) Li revenos hejmen je la plenluno.

- (96-en.) He will come home at the full moon.
- (97-eo.) Ili preterpasis ŝin je kelke de paŝoj.
- (97-en.) They pass over her of some steps.

Space relations and movement

Tida [2007] rightly argues with emphasis that "Esperanto has a lot of prepositions for space relations, and many of them do not exist in some languages". Most of them can be put in pairs, and I put apartly the complex adpositions $Prep + -n$. Table 5.11 shows space and time prepositions. The prepositions *en* and *el* can be rightly seen in pairs: *en* is borrowed from the German *in*, while *el* is borrowed from the German *aus*. Let me explain through a couple of examples.

- (98a-eo.) Liza iras en la ĝardeno.
- (98a-en.) Liza goes in the garden (she is always in).
- (98b-eo.) Liza iras en la ĝardenon.
- (98b-en.) Liza goes into the garden (she was out, now she is in).
- (98c-eo.) Liza iras ĝardenen.
- (98c-en.) Liza goes towards the garden.
- (99-eo.) Liza iras el la ĝardeno.
- (99-en.) Liza goes out of the garden (she was in, now she is out).

This subtle distinction with *en* and the nominative/accusative case (example 98) is rightly calqued from German – compare with the accusative circumstantial, section 5.3.6. The fact that *el* is calqued from German is testified by its use as prefix: *el-trink-i*, 'to drain', is clearly a calque of the German *aus-trink-en*. There are other moduli for *el*: *banejo el marmoro*, 'a pool made of marble (material)', *el surprizo li fuĝis*, 'because of surprise he escaped (cause)'. Finally, *el* indicates a precise individual in a set: *unu el ŝiaj amikoj*, 'one of her friends', or even *unuela amiko*, 'a friend (of somebody)'. I use the sign $\exists$ to indicate constructive existentials (see section 5.3.9 for details).

Nearness is expressed by two prepositions *ĉe* and *apud*. The difference is into the relation between the stative governor and the dependent.

- (100a-eo.) Liza estas ĉe (la) tablo.
- (100a-it.) Lisa è a tavola.

Table 5.11: The Esperanto space and time prepositions

Prep.	Gov.	Basic modulus	Additional m.
en	{O|I}	+[SPACE,INTERNAL]	+[TIME]
en + -n	{O|I}	+[SPACE,IN,TO]	+[TIME]
el	{O|I}	−[SPACE,INTERNAL]	+[TIME]
el	{O|I}	−[SPACE,INTERNAL]	+[CAUSE]
el	O	−[SPACE,INTERNAL]	+[MATERIAL]
el	O	−[SPACE,INTERNAL]	+[$\exists$]
ĉe	O	+[SPACE,NEAR],+[ACTIVE]	+[TIME]
ĉe + -n	O	+[SPACE,NEAR,-WARD] +[ACTIVE]	+[TIME,MOVE,END]
apud	O	−[SPACE,NEAR],−[ACTIVE]	+[TIME]
ĝis	{O|I}	+[SPACE,LIMIT]	+[MOVE],+[TIME]
ĝis	{O|I}	+[SPACE,LIMIT]	+[CHANGE],+[STATE]
ĝis	{O|I}	+[SPACE,LIMIT]	+[GRADIENT],+[LIMIT]
dum	I	+[TIME,CONTINUOUS]	
antaŭ	{O|I}	+[SPACE,+[BEFORE]	+[IMPORTANT],+[TIME]
malantaŭ	{O|I}	+[SPACE,−[BEFORE]	−[IMPORTANT],+[TIME]
ĉirkaŭ	I	+[SPACE,AROUND]	
ĉirkaŭ + -n	I	+[MOVE,AROUND]	
laŭ	{O|I}	+[MOVE,AROUND,DIR.]	
kontraŭ	{O|I}	+[MOVE,OPPOSITE,DIR.]	
post	{O|I}	+[SPACE,−[BEFORE]	−[IMPORTANT],
post	{O|I}	+[TIME,−[BEFORE]	−[IMPORTANT]
super	{O|I}	+[SPACE,OVER]	
super + -n	{O|I}	+[MOVE,OVER]	
sur	{O|I}	+[SPACE,ON]	
sur + -n	{O|I}	+[MOVE,ON]	
sub	{O|I}	+[SPACE,BELOW]	
sub + -n	{O|I}	+[MOVE,BELOW]	
inter	{O|I}	+[SPACE,TWO,MIDDLE]	+[TIME],+[$\exists$]
inter + -n	{O|I}	+[MOVE,TWO,MIDDLE]	+[TIME],+[$\exists$]
ekster	{O|I}	+[SPACE],−[INTERNAL]	±[$\exists$]
ekster + -n	{O|I}	+[MOVE],−[INTERNAL]	±[$\exists$]
preter	{O|I}	+[MOVE],−[INTERNAL], ±[AROUND]	
preter + -n	{O|I}	+[MOVE],−[INTERNAL], −[AROUND]	
tra	{O|I}	+[MOVE,INTERNAL]	+[TIME]
tra + -n	{O|I}	+[MOVE,INTERNAL]	−[AROUND]
trans	{O|I}	+[MOVE],−[INTERNAL]	+[TIME]
trans + -n	{O|I}	+[MOVE],−[INTERNAL]	−[AROUND]

- (100a-en.) Liza is at table.
- (100b-eo.) Liza estas apud (la) tablo.
- (100b-it.) Lisa è vicino alla tavola.
- (100b-en.) Liza is next to the table.

In example 100a Liza has an active relation with the table, while in 100b her relation is non-active (I am not saying passive!). The complex position *ĉe* with the accusative marker is identical to one of the uses of *al* already seen in Table 5.10.

The preposition *ĝis* is the last one proper Minus preposition, and it indicates the limits of a movement, which the governor does not go beyond (the limit is included in the movement).

- (101a-eo.) Ĝis kie ni iros hodiaŭ?
- (101a-en.) Where to will we go today?

Other uses are: *Mi ne senvestigu sin ĝis nudeco*, 'I won't undress her until nudity' (state change); *de homo ĝis bruto*, 'from man to beast (last step, limit in a gradient). Finally, *ĝis* can be used as a verbal lexeme to indicate 'good bye' and similar phrases – in this case it is *not* an adposition but part of the lexicon.

The preposition *dum* is a Plus adposition, and it is also used at a phrasal level: *tondroj bruadis dum la tuta nokto*, 'thunders crackled during the whole night (basic level)'; *restu apud mi, dum mi kun li ekstere parolas*, 'stay next to me, while I talk with him outside'.

According to the *Baza Radikaro Oficiala*, Esperanto has 15 morphemes with the pseudofinal *-aŭ*: some are adverbs, i.e., selective circumstantials, the others are prepositions.[15] Among them, there are five local prepositions: *antaŭ* and *malantaŭ* (example 102), *ĉirkaŭ* (examples 103-4), *laŭ* (example 105) and *kontraŭ* (example 106).[16] For the other morphemes with this termination, see section 5.3.8.

- (102a-eo.) Mi iras antaŭ vi.
- (102a-en.) I go before you.
- (102b-eo.) Mi iras malantaŭ vi.
- (102b-eo.) Mi iras post vi.
- (102bc-en.) I go after you.
- (103-eo.) La rabistoj sidis en rondo ĉirkaŭ la fajro.
- (103-en.) The bandits were sitting in circle around the fire.

- (104-eo.) Siajn brakojn ŝi metis ĉirkaŭ la kolon.
- (104-en.) Her arms she put around the neck.
- (105-eo.) Nun ni kuros supren laŭ la ŝtuparo.
- (105-en.) Now we will run upstairs following the staircase.
- (106-eo.) Unu figuro venis kontraŭ li.
- (106-en.) A figure was coming against him.
- (107-eo.) Post la vespermanĝo niaj fratoj eliris kun la gastoj.
- (107-en.) After the dinner our brothers went out with the guests.

The preposition *antaŭ* has two antonyms: *malantaŭ* and *post*. As a spatial preposition, *post* is a synonym of *malantaŭ* (example 102), while most often it is used as a time preposition (example 107). The prepositions *super* and *sur* are borrowed from the English *over* and *on*. The accusative can be applied in order to sign accusative (see example 108).

- (108a-eo.) Super ni brilas steloj.
- (108a-en.) Over us there are the brighting stars.
- (108b-eo.) La suno leviĝis super la teron.
- (108b-en.) The sun raised over the earth.
- (108c-eo.) Sur la tero kuŝas ŝtono.
- (108c-en.) On the earth there is a stone.
- (108d-eo.) Nenio nova sub la suno.
- (108d-en.) Nothing new under the sun.
- (108e-eo.) La muso kuris sub la liton.
- (108e-en.) The mouse run under the bed.

Also *inter*, *ekster* and *preter* are calqued on Latin, and their meaning should be quite clear through the following examples.

- (109a-eo.) Ni vojaĝis inter Romo kaj Pizo.
- (109a-en.) We travelled between Rome and Pisa.
- (109b-eo.) Vi elserĉu inter la tuta popolo homojn bravajn.

- (109b-en.) You search among the whole people until you find some brave persons.
- (110a-eo.) Mi staras ekster la pordo.
- (110a-en.) I am standing locking out.
- (110b-eo.) Li loĝas ekster la urbo.
- (110b-en.) He lives out of town.
- (111a-eo.) Ekster ŝi, ĉiuj venis.
- (111ac-en.) Except her, all came.
- (111b-eo.) Ekster la respondo letera, ni poste donos al vi ankaŭ respondon en nia gazeto.
- (111bd-en.) Beyond the answer by mail, we will afterwords give to you even an answer in our journal.
- (112a-eo.) Li pasis preter mi sen saluto.
- (112a-en.) He passed beyond me without greeting.
- (112b-eo.) La vojo kondukis preter preĝejon.
- (112b-en.) The way went well over {a|the} church.

The preposition *inter* can also express belonging, as in example 109: this existential is not constructive, i.e., it is not indicated clearly the individual in the set. Analogously, example 110 shows the existential uses of *ekster*, both additive and subtractive – also the Slavic preposition *krom* is used that way (see below). The use with the accusative is synonymic to preposition *el*. The preposition *preter* can be used with the accusative so to mark the fact that the governor is well away from the space of reference. The prepositions *tra* and *trans* are similar and different at the same time.

- (112a-eo.) Tra la fenestro.
- (112a-en.) Through the window.
- (112b-eo.) La sago iris tra lian koron.
- (112b-en.) The arrow went through his heart.
- (113a-eo.) Trans la limo.
- (113a-en.) Beyond the frontier.
- (114a-eo.) Ŝi kuris tra barilon.

Table 5.12: The Esperanto non spatial *-aŭ* prepositions

Prep.	**Governor**	**Basic modulus**	**Additional moduli**
anstataŭ	{O|I}	+[NEGATION],+[SUCCESS]	
malgraŭ	{O|I}	+[NEGATION],−[SUCCESS]	
kvazaŭ	{O|I}	+[NEGATION],−[SUCCESS,MILD]	
preskaŭ	{O|I}	+[NEGATION,MILD],−[SUCCESS]	

- (114a-en.) She run through a barrel.
- (114b-eo.) Ŝi kuris trans barilon.
- (114b-en.) She run beyond a barrel.

The preposition *tra* expresses a movement through the space, from one border to the other, while *trans* expresses a movement in order to reach the other side, without going into the space (see example 114 in particular).

The non-spatial *-aŭ* prepositions

Table 5.12 shows non spatial prepositions which finish in *-aŭ*. The preposition *anstataŭ* has no direct equivalent in English, which use complex adpositions like 'instead of, in lieu of, in place of' (example 115). The preposition *malgraŭ*, borrowed from French, is complementary: the substitution did not succeed (example 116).

- (115-eo.) Anstataŭ kafo li donis al mi teon.
- (115-en.) In spite of coffee he gave to me a tea.
- (116-eo.) Malgraŭ ĉio mi ne povas.
- (116-en.) In spite of everything I cannot.
- (117a-eo.) Karlo estis kvazaŭ pikita per ponardo.
- (117a-en.) Charles was hit by a dagger-like.
- (117b-eo.) Karlo estis preskaŭ pikita per ponardo.
- (117b-en.) Charles was almost hit by a dagger.

The nuance between *kvazaŭ* and *preskaŭ* is very subtle, but important, as clarified by example 117: construal 117a shows that Charles was hit by something that resembled a dagger but it is not, while in construal 117b Charles was eventually not hit.

Table 5.13: Other prepositions

Prep.	Governor	Basic modulus	Additional moduli
da	O, +[MEASURE]	+[QUANTITY]	
krom	{O|I}	+[ABSTRACT], −[INTERNAL], ±[$\exists$]	
por	I	+[GOAL]	+[N_e]
pro	I	+[CAUSE]	+[N_a|N_x]
pri	I	+[TOPIC]	
per	{O|I}	+[TOOL,RESOURCE]	+[N_a|N_x]
kun	{O|I}	+[COMPANY]	+[TIME]
sen	{O|I}	−[PRESENT]	

The other prepositions

The preposition *da* is very specific. In fact, it is used only within stative groups indicating measure and quantity (typically, it is used with *-om* correlatives: see section 5.3.9).

- (118a-eo.) Unu kilogramo da pano.
- (118a-en.) A kilogram of bread.
- (118b-eo.) Skatolo da rizo.
- (118b-en.) A box (full) of rice.
- (118c-eo.) Multe da problemoj.
- (118c-en.) A lot of problems.

The preposition *krom* is calqued from Russian. Its basic use is very similar to the secondary use of *ekster*, save for the fact that *krom* is not spatial in nature.

- (111c-eo.) Krom ŝi, ĉiuj venis.
- (111d-eo.) Krom la respondo letera, ni poste donos al vi ankaŭ respondon en nia gazeto.

The prepositions *por*, *pro*, *pri* and *per* look very similar but they are very different. The preposition *por* expresses a goal of the action, and sometimes an Experiencer (example 119b), while the preposition *pro* express a cause (example 120).

- (119a-eo.) Mi ne havas tempon por tio.
- (119a-en.) I don't have time for that.
- (119b-eo.) Tiuj bonaj homoj certe trovos por mi laboron.
- (119b-en.) Those good persons will certainly find a job for me.
- (120-eo.) Ŝiaj dentoj frapadis pro malvarmo.
- (120-en.) Her teeth were chattering because of the cold.
- (121-eo.) Mi parolas pri leono.
- (121-en.) I'm talking about a lion.
- (122a-eo.) Per hakilo mi hakas, per segilo mi segas.
- (122a-en.) With a hack I hack, with a saw I saw.
- (122b-eo.) Ŝi ne havis piedojn, la korpo finiĝis per fiŝa vosto.
- (122b-en.) She didn't have feet, the body finished with a fishtail.
- (123a-eo.) Resti kun leono estas danĝere.
- (123a-en.) To stay with a lion is dangerous.
- (123b-eo.) Leviĝu kun la suno.
- (123b-en.) Wake up with the sun.
- (124a-eo.) Ŝi iras sen celo.
- (124a-en.) She goes without goals.

The preposition *pri* introduces the topic, hence it is always a circumstantial (example 121), while preposition *per* indicates the resources to achieve a result or a goal (example 122).

The symmetrical preposition *kun* indicates essentially company (example 123a), and sometimes it indicates time (example 123b). Finally, the preposition *sen*, which is Romance (e.g., French *sans*, Italian *senza*), corresponds to the English grammaticalised compound *with-out*. There are no other prepositions beyond the one presented above.

Table 5.14: The structure of correlatives[Szerdahelyi, 1989, extract]

Esperanto	English	Latin	French	Italian	Spanish
kiu, kia	what	qualis	quel	che	qué
kiu, kia	which	qui	lequel de	quale	cuál
tiu	that	ille, iste	ce... lá	quel	aquel
tiu ĉi	this	hic	ce... ci	questo	ese
iu	some	aliqui(s)	quelque	qualcuno	alguno
iu ajn	any	ullus	quelquonque	qualunque	cualquiera
ĉiu	each	quisque	chaque	ciascuno	cada uno
ĉiuj	all	omnes	tous	tutti	todos
neniu	no	nullus	aucun, nul	nessuno	ninguno

5.3.9 Correlatives

Correlatives are special morphemes used to introduce questions and relative clauses and to give appropriate answers. They have a particular status also in implementation. In adgrams, questions and so-called "relative" clauses are considered collocational phenomena, where the order of morphemes is changed along precise rules, analogously to what seen about mirroring in section 2.9.3. This means, that they are built over the neutral, standard word order and then "rotated" appropriately. They are treated together because the phenomena they activate are structurally similar: in particular, they can augment the valence.

Questions can be divided along the possible answers, as yes/no questions or 'other', more complex: This is cross-linguistically valid. Furthermore, at least in Indo-European languages, the *wh*-questions are "relative" clauses that can be nested. The connectors form a specific set: each member of this set I call **correlative**. More or less, Indo-European languages show a morphological paradigm underneath correlatives: in English it can be found through the pseudoprefixes *wh-* and *th-*.[17] Table 5.14 shows the structure of correlatives in some Indo-European languages in comparison. For example, in French some correlatives have *quel-* as the pro forma and *-que* as the forma, while in Italian the same pro forma is *qual-*. By the way, English – as other NLs – can express some correlatives only invoking complex adpositional patterns, as the its correlative system is less regular.[18] A correlative is a grammaticalised morphemic combination between a pseudoprefix, which I call **pro forma**, and a pseudosuffix, which I call **forma**.

The pro forma is the coolest landmark possible, save obviously for zero morphemes, i.e., pure syntax. Pro formas can be used to introduce question or relative clauses or to specify abstract informations belonging to the answer. I call the clause pro formas **C pro forma** ($\mathcal{C}$), while the possible answers **D pro formas** ($\mathcal{D}$). The set of $\mathcal{D}$ has all and only the following members:

- there is no desired element in the set ($\nexists$-ProForma);
- all elements in the set satisfy the question ($\forall$-ProForma);
- there exists an element ($\exists$-ProForma);
- there are elements in the set not better defined ($\in$-ProForma).

Formas give the grammar character to the correlatives they belong. Of course, verbal correlatives are simply impossible without an additional transferer, so correlatives can be circumstantial, stative or adjunctive.[19] In Esperanto, it is noteworthy that some correlatives can accept the pluraliser *-j* and/or the accusative signature *-n* (example 126), while others are invariable (example 125).

- (125-eo.) Mi diras ion, (tial) kial vi povu aŭdi.
- (125-en.) I say something (so that) for the reason that you can hear.
- (126-eo.) Mi diras (tiun), kiun vi povu aŭdi.
- (126-en.) I say something (that) you can hear.

The cognitive functions of $\mathcal{C}$ and $\mathcal{D}$ pro formas are different, therefore they are represented in two different ways in the adpositional grammars.

Correlative clauses

In Esperanto, the pro forma for questions is always *ki-*.[20] The function of the $\mathcal{C}$ pro forma is to open a clause. I call **clause** a pseudophrase, i.e., a syntactic object built as a phrase but 'living' only in symbiosis of another phrase, which governs the clause itself. I call it the **clause governor**. If it is not expressed, that clause is a **question**. This can happen even without explicit question marks: Esperanto has an explicit morpheme to indicate questions (see below, section 5.3.10 for details).

Each clause is a masked circumstantial or adjunctive. Let me bring you beyond this veil through another example.

- (127a-eo.) Kiam pluvas, mi dormas.
- (127a-eo.) Pluve, mi dormas.
- (127ab-en.) When it's raining, I sleep.

Table 5.15: Adjunctive correlatives (a-corr-group)

pro forma / forma	+[QUESTION, RELATIVE]	∈	∃	∀	∄
	ki-	*ti-*	*i-*	*ĉi-*	*neni-*
+[QUALITY]	ki-a(-jn)	ti-a(-jn)	i-a(-jn)	ĉia(-jn)	neni-a(-jn)
+[POSSESS]	ki-es	ti-es	i-es	ĉi-es	neni-es
+[DEFINITIVE][21]	ki-u(-jn)	ti-u(-jn)	i-u(-jn)	ĉi-u(-jn)	neni-u(-n)
−[DEFINITIVE]	ki-o(-n)	ti-o(-n)	i-o(-n)	ĉi-o(-n)	neni-o(-n)

Table 5.16: Circumstantial correlatives (e-corr-group)

pro forma / forma	+[QUESTION, RELATIVE]	∈	∃	∀	∄
	ki-	*ti-*	*i-*	*ĉi-*	*neni-*
+[TIME]	ki-am	ti-am	i-am	ĉiam	neni-am
+[PLACE]	ki-e	ti-e	i-e	ĉi-e	neni-e
+[MOVE]	ki-en	ti-en	i-en	ĉi-en	neni-en
+[MANNER]	ki-el	ti-el	i-el	ĉi-el	neni-el
+[CAUSE]	ki-el	ti-el	i-el	ĉi-el	neni-el

- (127c-eo.) Kiom mi ŝatas dormi kiam pluvas!
- (127d-eo.) Kiomŝata estas dormado kiam pluvas!
- (127cd-en.) How much I like to sleep when it's raining.

The flexibility of Esperanto helps to unmask this syntactic reality: example 127ab shows how the relative clause (*kiam pluvas*) is a circumstantial (*pluve*), while example 128cd shows how the relative clause (*kiom mi ŝatas*) is an adjunctive (*kiomŝata*) – remember that the infinitive is a masquerade performed by a verb, transferred into a stative.

The grammar characterisation of the correlative is an information which is crystallised a priori in the correlative, because of grammaticalisation. The following tables give the correspondence between adtypes and formas in correlatives. Table 5.15 shows adjunctive correlatives, while Table 5.16 shows circumstantial correlatives. Note that these last are all invariable.[22] Let me recall and expand example 32, in order to explain this tricky point of the Esperanto grammar, inherited from its Indo-European source languages.[23]

- (32a-eo.) Kion fumas Karlo?

- (32a'-eo.) Kion Karlo fumas?
- (32a-en.) What does Charles smoke? (e.g., cigars? cigarettes?)
- (32b-eo.) Kiun fumas Karlo?
- (32b'-eo.) Kiun Karlo fumas?
- (32b-en.) Which one does Charles smoke?
- (32c-eo.) Kiu(j)n cigaro(j)n fumas Karlo?
- (32c'-eo.) Kiu(j)n cigaro(j)n Karlo fumas?
- (32c-en.) Which cigar(s) does Charles smoke?
- (32d-eo.) Karlo fumas kian cigaron li ŝatas.
- (32d-en.) Charles smokes the type of cigars he likes.
- (32e-eo.) Kian cigaron fumas Karlo? Toskanojn.
- (32e-en.) Which type of cigar(s) does Charles smoke? Tuscans.
- (32f-eo.) Por kiu (motivo) Karlo fumas (cigarojn)?
- (32f'-eo.) Kial Karlo fumas (cigarojn)?
- (32f-en.) Why does Charles smoke (cigars)?

It is noteworthy that the accusative can be in a marked position, as the word order here can be OVS instead of SVO. The semantic opposition between the formas in *-u* and in *-o* is explained in examples 32a and 32b. Of course, if something is not defined it will not be pluralised at all: the forma in *-o* cannot be pluralised by definition. In contrast, if something is definite it can be explicitely defined furtherly; in other words, the forma *-u* can specify a stative, which can be empty.[24] Note that it is always possible to apply a suppletive strategy instead of a grammaticalised one (example 32f). Finally, note that the series made with the forma *-es* is a saxon genitive.[25] The adtree should be built accordingly (Figure 5.6). In practice, the degree of grammaticalisation in $\mathcal{C}$**-correlative**, i.e., correlatives for questions, is so strong that there is no need to split pro formas and formas. In fact, I realised that in a preliminary version of this dissertation adtrees where more complex than necessary because of this fixed, freezed adpositional tree chunck inserted in – this structure has no influence on the rest of the adtree. For this reason, in the formal model presented in chapter 7 $\mathcal{C}$-correlatives are considered as a single, cristallysed token, and therefore the final adtree will become far more easy to manipulate (Figure 5.7). Of course, the adtype is Slash, as it grammaticalisation here is very, very strong. Figure 5.8 explains

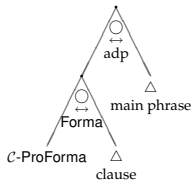


Figure 5.6: The theoretical abstract adtree of $\mathcal{C}$-correlatives

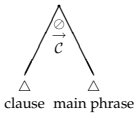


Figure 5.7: The reasonable abstract adtree of $\mathcal{C}$-correlatives

correlatives and at the same time it gives a glance of how sentences are applied, essentially the same mechanism of phrasal juctive (see 5.4). In fact, the presence of the question mark changes drammatically the abstract tree Figure 5.7. But also example 32f respects it, as the clause is absent, or, better said, it is void (□). Note that *kian* is singular, while *Toskanojn* is plural.[26]

Determiner correlatives

Figure 5.9 retains shows that the abstract tree of $\mathcal{D}$-correlatives is very similar to the first, theoretical adtree of $\mathcal{C}$-correlatives (Figure 5.6). Let me explain it expanding example 32 furtherly.

- (32g-eo.) Karlo fumas ĉies cigarojn.
- (32g-en.) Charles smokes the cigars which belong to everyone.

Figure 5.10 shows that determiner correlative, as the name suggests, are a subclass of determiners. For a full explanation, see section 5.8. What is important to note since now is the fact that if the leaf of *cigar-* is void, the adtree does not change: the leaf will be simply boxed (□).

5.3.10 Junctives

Any (Q)NL presents complex sentences, i.e., sentences with more than one phrases. In this section I present briefly the phrasal adpositions – i.e., junctives – of Esperanto. It is noteworthy that the formal model built on the sentence level – and even more at a text level – is already general, i.e., almost independent from Esperanto.

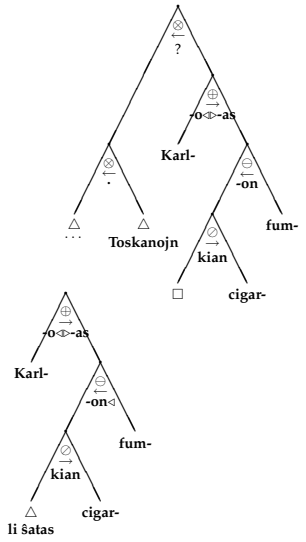


Figure 5.8: $\mathcal{C}$-correlatives in action (example 32ef).

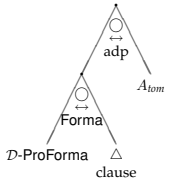


Figure 5.9: The abstract adtree of $\mathcal{D}$-correlatives

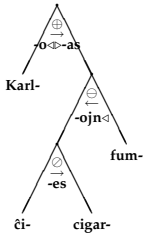


Figure 5.10: Determiner correlatives in action

In contrast with basic adpositions, *phrasal adpositions can have moduli, i.e., they show some lexical value*, e.g., the English junctive *while* is marked with the modulus +[TIME], while the junctive *because* is marked with the modulus +[CAUSE].[27] As explained in section 2.10, the phrasal adspace is only retroapplicative, i.e., junctives can be either non-dimensional (Slash) or dimensional (Times). Let me recall and translate example 43:

- (43a-eo.) △(Alfredo povas pagi), ĉar li estas riĉa.
- (43a-en.) △(Al can pay), because he is rich.
- (43b-eo.) Alfredo povas pagi, do △(li estas riĉa).
- (43b-en.) Al is rich, {hence|therefore} △(he can pay).

These are the most used phrasal adpositions in Esperanto (translations are only indicative):

- **Slash**: *sed*, 'but', *tamen* 'nevertheless', *kvankam* 'however', *se* 'if', *ke*, 'that', *ki-* 'wh-', *do*, 'so';
- **Times**: *kun*, 'with', *dum*, 'while', *ĉar*, 'as', *aŭ*, 'or', *kaj*, 'and', *nek*, 'nor'.

Adpositions can be switched between the two levels only along the applicativity axis: e.g., *dum*, 'while', is a Plus adposition as a preposition,

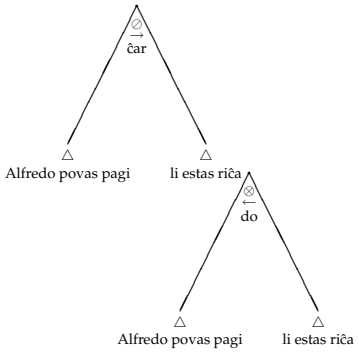


Figure 5.11: How to build adtrees with phrasal adpositions (example).

while it is a Times adposition as a junctive.[28] Figure 5.11 shows how to join adtrees via phrasal adpositions – branches are longer for readability. In the following, I will clarify the structure of adpositions, not pretending to be exhaustive. I give only two examples of two Slash junctives which are very common in use: *se*, 'if', and *ke*, 'that'. Let me use two examples.

- (129-eo.) Paŭlo scias, ke Liza studas.
- (129-en.) Paul knows that Liza is studying.
- (130-eo.) Paŭlo ne scias se Liza studas.
- (130-en.) Paul doesn't know if Liza is studying.

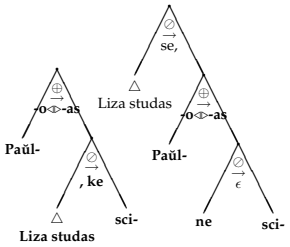


Figure 5.12: The adtrees of *Paŭlo (ne) scias...* (129-130).

Figure 5.12 (left) shows the use of *, ke*, which introduces even a whole sentence with the Patient – syntactically, this is often considered rightly as an accusative.

Binary questions and negation

Esperanto borrowed from the Polish *czy* the morpheme *ĉu* in order to introduce yes/no questions, which is applied in the adtree as an external circumstantial (void actant $X_{1,\ldots,n}$), without touching the word order – see Figure 5.13 for the corresponding adtrees.

- (32-eo.) Karlo fumas en la vojtruon.
- (32-en.) Charles smokes in the manhole cover.
- (32h-eo.) Ĉu Karlo fumas en la vojtruon? (Jes,ne)

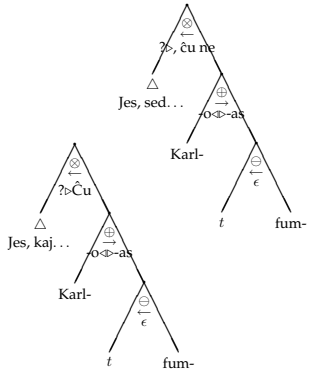


Figure 5.13: Esperanto question clauses (example 32).

- (32h-en.) Is Charles smoking in the manhole cover? (Yes,no)
- (32i-eo.) Ĉu Karlo ne fumas en la vojtruo? (Jes ja, ne)
- (32i-en.) Is Charles not smoking in the manhole cover? (?Yes,no)
- (32j-eo.) Karlo fumas en la vojtruo, ĉu ne?
- (32j-en.) Charles is smoking in the manhole cover, isn't he?

Example 32hj shows that Esperanto borrowed the German word *doch* – which eventually became *ja* – in order to solve ambiguities in negative questions. The combination *ĉu ne* is an invariable idiomatic expression for *isn't he* and similar expressions.

Double junctives

Some junctives can be double, like the English *either... or*, *both... and*. Let us assume that the first adposition is the left member while the second adposition is the right member. The adtree is built in the same way even if the left member is omitted, as shown by the following example 128:

- (128-eo.) Kaj Liza kaj Paŭlo legas libron.
- (128-en.) Both Liza and Paul read a book.

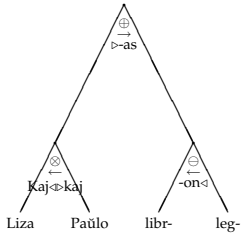


Figure 5.14: The adtree of *Kaj Liza kaj Paŭlo legas libron* (128).

Figure 5.14 shows how the adtrees are built in these cases. In the third part all rules so to treat these phenomena will be given.

5.4 Sentence delimiters

Last but not least, punctuation can be used as a junctive. This fact is cross-linguistical, otherwise no text would be possible. A text is a linear structure,

Table 5.17: The five levels of analysis in Esperanto

Level	Elements	Junctors	Examples
Text level	sentences	□	□
Sentence level	phrases	junctives	{?|.|!. . . }
Phrase level	clauses	adpositions	{de|al|pro. . . }
Clause level	groups	$\mathcal{C}$-corr.	{kiu|kiam|kien. . . }
Group level	subgroups	adpositions	{ϵ|-j|-n. . . }

i.e., it is clear what is before, and what is afterwards, so the adtree at the text level is trivial. I have already given an example in Figure 1, i.e., Terence's quotation that opens this dissertation. Sentence delimiters can be Slash as the adtype, as in that example, or Times, as full stops or question marks, exactly as junctives. Let me explain this point through a very simple text.

- (128a-eo.) Ĉu Liza legas libron? Ne, ŝi studas. Fakte, ŝi lacas.
- (128a-en.) Does Liza read a book? No, she is studying. In fact, she is sleepy.

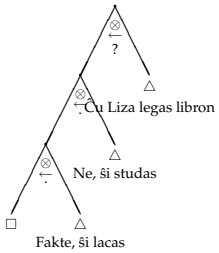


Figure 5.15: The adtree of a short text (example 128a, unexploded).

Figure 5.15 shows how the adtree of a whole text is built (I will come back later in this topic, in chapter 8, devoted to MT). What I want to underline here is the evidence of five levels of analysis in Esperanto. Table 5.17 shows the five levels of analysis in NLs; a **junctor** is simply a hyperonym of junctives, prepositions, $\mathcal{C}$-correlatives. Of course, some levels, most notably the clause levels, can be not operative in actual use.[29] This ends the description of the Esperanto adspace: now the reader is invited to turn his or her attention to the Esperanto lexicon.

5.5 The Esperanto lexicon

The Esperanto lexicon is made of calques from the lexemes belonging to its source languages: e.g., *krajono* is from the French, *biero* is from German, *sed* is from Latin, *ĉu* is from Polish. Interestingly, Zamenhof preferred to choose a *Lautbild*, i.e., a phonematic writing system, over a *Schriftbild*, i.e., an etymological writing system. For instance, lexemes borrowed from German NLs are calqued as the best compromise between English and German: *strato* is a compromise between *street* and *Straß*; *futbalo* is a compromise between *football* and *Fußball*, and so on. Lexemes borrowed from Latin, Greek or Romance languages are already adapted to the Slavic-like phonological rules: e.g., the Latin *aqua* and *quamquam* become in Esperanto respectively *akvo* and *kvankam*.

In 1974 the Academy of Esperanto has published the *Baza Radikaro Oficiala*, the basic dictionary of the language, now available both in the official web site of the Academy and as a Wikisource.[30] That work is the main reference of this section. However, there is one important point of difference between adgrams and the approach of the Academy insofar. In the adpositional grammar representation of Esperanto, the fundamental grammar characters are four, instead of three, as in the document of the Academy. I argue that the character of words like *ĵus*, 'just', or *morgaŭ*, 'tomorrow', is clearly circumstantial (E), and I regard this as a matter of fact, as they modify verbs: otherwise, the whole system presented here collapses.

5.5.1 Selection and redundancy

Some Esperanto lexemes are selective, i.e., nouns, verbs, adjuncts, adverbs, while others are clear only if collocated. For example, *elefant* and *polic* are nouns, *honest* is an adjective while *kis* is a verb: in the sentence *honest-a polic-o kis-as elefant-o-n*, 'the honest police kisses an elephant' the signatures of the fundamental grammar characters are redundant, as the sentence in an "Esperanto sine flexione" à la Giuseppe Peano *honest polic kis elefant* seems to be equivalent.[31] This redundancy may appear pedantic, especially to English readers, as in the English language most lexemes are *not* selective; however, the Esperanto flexibility of its final system is indoubitably a source of richness: *kis-a honest-o polic-as elefant-e*, 'the honesty concerning kisses acts as police like elephants do'.[32]

In some cases, the grammar character seems to be more relaxed: *muzik-* and *bicikl-* are used indifferently as stative, verbal, adjectival and circumstantial. Nonetheless, the structure of the dictionary, as presented in chapter 3, request some additional information, e.g., valence and the actants, so that lexemes can be classified accordingly. In other cases, some transfers are not clear semantically, so in practice they are avoided: the verbal use of the selective noun *libr*, 'book', is never used, even if in English 'to book' has

a very precise meaning: this is a further proof of the linguistic autonomy of Esperanto. In the dictionary, the transfer O>I of the lexeme *libr* will be simply blocked (i.e., marked with a $\perp$ in the dictionary). Moreover, even in Esperanto combinability of affixes is not free, but governed by precise rules, according to the grammar character of the lexeme and sometimes to the moduli that are into the lexemes (see details below).

5.5.2 Rules for compounding

Esperanto forms compounds following the Germanic model – more precisely as German, Yiddish and English, which are its Germanic source languages.[33] Compounding is perhaps the most highly form of grammaticalisation in Esperanto, as the meaning of the relation of the participants, i.e., their adposition, is completely dependent on the lexemes themselves. For instance, *vapor-ŝip-o* is 'steamship', while *ŝip-vapor-o* is the steam of a ship. I call the first element the **lexical specification**, while the second element the **lexical atom**. In a compound, there can be at best one lexical specification and atom at a time.[34] Affixes can be applied either, but they follow their own rules, described below.

The grammar character of the atom is crucial. First, it influences how the grammar character of the specification is transferred, i.e., the rules of the compounds depends on the grammar character of the atom. Hence, there will be four compound rules: stative, adjunctive, verbal and circumstantial. In practice, circumstantial compounds are not attested, hence the rules are three.[35] Furthermore, circumstantials do not participate as specifications, and this permits a further simplification of the rules themselves. Let us consider the adpositions for compounding special zero adpositions which acts as additional transferers. But the grammar character of the atom is very important for another crucial reason: *the set of affixes that can be applied to the compound depends on the grammar character of the atom*. In fact, there are some **general affixes**, i.e., affixes that can be applied to any lexeme, regardless of its grammar character, but there are also **stative affixes**, i.e., affixes that can be applied only to stative lexemes as governors, and **verbal affixes**, i.e., affixes that can be applied only to verbal lexemes as governors. There are not adjunctive or circumstantial affixes, i.e., specific only for adjunctive or circumstantial lexemes.[36] Table 5.18 shows the relation between grammar character and affixes. The lexeme *salt-*, 'to jump', is verbal, so the verbal prefix *ek-* can be applied. In contrast, the lexeme *knab-*, 'boy', is stative, so the applying of *ek-* gives a non-sense. Analogously, the lexeme *verd-*, 'green', is adjunctive, so the suffix *-ul-*, which gives animation to the inanimate, can be applied safely. Note that also *tabl-*, 'table', is inanimate, but stative, so *-ul-* cannot be applied either. Note that the final grammar character of the word is completely uninfluent in this respect.

At a first glance this linguistic fact seems to be a severe limit to the

Table 5.18: The grammar character and the affixes (specimen)

Analysed word	Gram. char. of the lexeme	English translation	Italian translation
ek-salt-o	I	a leap	un balzo
*ek-knab-o	O	?	?
verd-ul-a	A	relating a green (man, politician...)	di un verde
*tabl-ul-a	O	?	?

Table 5.19: The grammar character and the special affixes (specimen)

Analysed word	Modified gram. char.	English translation	Italian translation
sur-tabl-ig-o	O>I	the laying of the table	l'apparecchiatura
*sur-tabl-o	O	?	?
sur-met-e	I	with the act of putting on (a dressing)	con il mettersi su, il vestirsi
(ŝi) ek-ruĝ-iĝ-as	A>I	she started to turn red	lei diventò rossa
(ŝi) *ek-ruĝ-as	A	?	?
(ĝi) ek-trink-iĝ-os	A>I	it will be start to be drinked (by someone)	sarà iniziato ad essere bevuto (da qualcuno)

freedom of Esperanto morphology. Nonetheless, there is a couple of special lexemes, which modify the valence of the word, which also act as inline verbal transferers. Table 5.19 shows the relation between the grammar character and the affixes. For example, the prefix *sur-* is derived from the homonymous preposition, and it is verbal in nature: in fact, it cannot be applied to the stative lexeme *tabl* unless it is verbified by the special suffix *-ig-*. Of course, these rules can be applied in the same way to compounds, after the substitution from 'grammar character of the lexeme' to 'grammar character of the atom'.[37] Now, let us see the detailed rules for compounding. Table 5.20 shows the rule of stative compounds. Basically, the final grammar character of the specification transferred by the O-compound adposition is always stative, regardless of the original grammar character: *paper(o)korbo*, 'a paper basket', *skribmaŝino*, 'a typewriter', *beldi(in)o*, 'a

Table 5.20: The rule of stative compounding (O-compound)

Specification	Atom
O>O	O
paper(o)-	*-korbo*
I>O	O
skrib-	*-maŝino*
A>O	O
bel-	*-di(in)o*

Table 5.21: The rule of adjunctive compounding (A-compound)

Specification	Atom
O>O	A
ŝton-	*-kolora*
A>O	A
bel-	*-riĉa*
I>O	A
servo-	*-preta*

God(dess) of Beauty'. Table 5.21 shows the rule of adjunctive compounds. This rule is analogue to the stative one: the final character of specification is always stative. In fact, the rare compound *neĝe-blanka*, 'snow white', found by Gledhill [1998], cannot be written as *neĝblanka*, because it would mean 'white because of the snow'. The examples are: *ŝtonkolora*, 'stone coloured', *belriĉa*, 'rich, full of beauty', *servopreta*, 'ready for service'.

Table 5.22 shows the rule of verbal compounds, which are more nuanced by the grammar character of the specification. Stative and verbal lexemes becomes modifiers of the atom, i.e., circumstantials: *pafmurdi* is the grammaticalisation of *murdi pafe*, 'to murder somebody through a shooting', while *finfari* is 'to do finally'. Analogously, *matenmanĝi* is the grammaticalisation of *manĝi matene*, 'to eat in the morning', that eventually is 'to have breakfast', while *fingromontri* is the grammaticalisation of *montri fingre*, 'to show through a finger'; *voĉdoni* means 'to vote', which is the grammaticalisation of to give (*doni*) with the voice (*voĉe*).

In contrast, adjunctive lexemes retains their grammar character, which is always considered predicatively: *ruĝpentri* is 'to paint something or somebody red', while *plenŝtopi* is exactly the German expression *voll stopfen*, 'to fill (something at a full level)'.

Table 5.22: The rule of verbal compounding (I-compound)

Specification	Atom
I>E	I
paf-	*-murdi*
fin-	*-fari*
O>E	I
maten-	*-manĝi*
fingro-	*-montri*
voĉ-	*-doni*
A>A(⊖)	I
ruĝ-	*-pentri*
plen-	*-ŝtopi*

Table 5.23: The Esperanto adposition spec-atom (specimen)

Spec	Atom	Translation	Added modulus
vapor-	-ŝipo	a steamship	+[ENERGY]
petrol-	-ŝipo	an oil tanker	+[ENERGY]
aer-	-ŝipo	an aircraft	+[SPACE]
kiras-	-ŝipo	an ironclad	+[MATTER]
batal-	-ŝipo	a battleship	+[GOAL]
karb-	-ŝipo	a collier	+[LOAD]

Lexicalisation

There are exceptions to the rules just presented. Normally these exceptions are due to the influence of the Greek or Latin substrata into the source languages of Esperanto. For example, *hom-manĝi* is not *manĝi home*, 'to eat (I) as a man (O>E)', but 'to act as a cannibal', i.e., *manĝi homo(j)n*, 'to eat (I) men (O>O)', as in the Greek-derived word *anthropophagy*. Analogously, *grand-anim-o* is not '?a soul of the bigness', but *magnanimous*, even if *anim-grand-a*, which follows the rules, is possible, but almost not used [Cherpillod, 2003]. All these cases should be treated as lexicalisation. Let me explain through the paradigm of stative compounds whose atom is *ŝipo* (ship).

Table 5.23 [Butler, 1947, elaboration], shows clearly that the adposition spec-atom depends *both* on the specification and the atom. So, what to do? Let us suppose that *vaporŝipo* is the syntactic subject S in a phrase (Figure 5.16). Theoretically, its adtree of words is the one described on the left. The problem is that most probably the machine translation engine would translate in English as 'a ship moved by steam' or something similar,

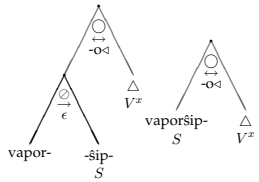


Figure 5.16: Adpositional trees of the word *vaporŝipo*.

instead of 'steamship'. This happens because *adtrees do not encode the encyclopædic knowledge necessary to solve this kind of ambiguities*: the only way to avoid it is consider *vaporŝipo* a **lexicalisation**, i.e., a compound which has been crystallised in the semantic space of Esperanto. Of course, it is highly questionable which compounds are lexicalised and which not: only a collocation analysis of monolingual corpora can give the answer. In terms of adtrees, the compound would be not analysed but on the contrary considered as a single lexeme (Figure 5.16, right).

Analogously, a *paf-il-eg-o* is not 'a (generic) big gun', but a cannon – although *kanon-o* does exist too; *mal-san-ul-ej-o* is not 'a (generic) place for sick people', but a 'hospital' – again, *hospital-o* does exist. The word *fer-voj-o* is not 'a way of iron', but a 'railway'. If *fervoj-* would not be considered as a lexicalisation how the word *fervoj-ist-o* can be understood as 'railwayperson'?[38]

Sometimes segmentation of compounds is not clear. The word *konkludo* is *konklud-o*, 'conclusion', or *konk-lud-o*, 'a play (*ludo* with shell *konkoj*'? Even the very name of the language, *esperanto*, can be analysed both as *esperant-o*, 'the language Esperanto', or *esper-ant-o*, 'one who hopes'. Esperanto is a living language, not a perfect language, with its own ambiguities: these cases will be treated in the implementation as different possible adtrees pertaining the same family. This solution is to be considered as a pragmatical one: paraphrasing Winston Churchill, it is not a perfect solution, but we do not have anything better.

5.5.3 Esperanto lexemes

In the dictionary, lexemes are stored taking into account every possible transfer from their basic grammar character. Moreover, valence and the actant schema – e.g., if the syntactic subject is an Agent or a Patient – are informations stored too, even if they are not always used. The conventions followed for the dictionary were already explained in chapter 3. Every lexeme should have a fundamental final attached so to clarify the final grammar character of the group. In the following two subsections special cases

Table 5.24: Esperanto adverbs and transference (specimen)

Adverb	Signature	Transfer	English translation	Italian translation
hodiaŭ	∅	E	today	oggi
hodiaŭ	-o	E>O	today	l'oggi
hodiaŭ	-a	E>A	today's	odierno
hodiaŭ	-e	E>E	nowadays	oggigiorno

will be clarified.

Adverbs

There is a special class of adverbs in Esperanto, i.e., lexemes that are circumstantial by selection, and Minus as adtypes. Unlike any other lexeme, they are stand-alone, i.e., no final character is attached: *ne*, 'no', *eĉ*, 'even', *jam*, 'yet', *ĵus*, 'just', *tro*, 'too much', are some examples.

A subclass of them is the series of adverbs in *-aŭ*, which should not be confused with the prepositions in *-aŭ*, already seen in Table 5.12. This series is limited and it is not etymologically homogeneous: *hieraŭ*, 'yesterday' (from French), *hodiaŭ*, 'today' (from French), *morgaŭ*, 'tomorrow' (from German), *adiaŭ*, 'good bye' (from French), *baldaŭ*, 'soon' (from German), *apenaŭ*, 'barely' (from French), *preskaŭ*, 'almost' (from French).

Of course, it is still possible to perform transference even for adverbs. In this case the meaning changes, i.e., there are some moduli to be added.

Extreme lexicalisation

Esperanto is a living language, and hence unplanned linguistic phenomena can emerge. One of the most interesting and compelling phenomenon is the lexicalisation of adpositions (sic!).

- (131-eo.) Ni kunigas oson al peranto.
- (131-en.) We put together the future to the mediator.

Words like *kun-ig-as* and *per-ant-o* are neither strange nor so infrequent, and their lexical base is respectively the lexicalisation of the prepositions *kun-* and *per-*. An even most extreme lexicalisation is the verbal signature of the future *os-*: this is experimental and not used, but not impossible in principle. In principle, each adposition can be lexicalised directly in Esperanto: e.g., *ĉui*, 'to put yes/no questions', which is rare but attested in my personal experience of Esperantophone. However, adgrams can treat even these extreme cases. In fact, these lexicalisations should be treated as so, i.e., they

are different entries in the dictionary, compared to prepositions they derive from.

5.5.4 Stative lexemes

In the lexicon, stative lexemes can be orderd on the abstractiveness axis, i.e., along hyponymy and hyperonymy in order to find the basic level – onomasiological salience [Taylor, 2002]. This is cross-linguistically valid. In Esperanto there are other dimensions along which stative lexemes can be orderd: animateness, gender, and number:

- Inanimate neutral: *tabl-*, 'table', *aparat-* 'tool', *bibliotek-*, 'library'.
- Animate neutral: *hom-*, 'human being', *fung-*, 'mushroom', *arb-*, 'tree'.
- Collective neutral: *popol-*, 'people', *eklezi-*, 'Church (not 'church'!)'.
- Animate masculine: *vir-*, 'man', *knab-*, 'boy'.
- Animate feminine: *meger-*, 'hag', *matron-*, 'matron', *putino*, 'bitch'.

The appropriate category can be verified with the application of the appropriate transferees – see section 5.6.6 below.

5.5.5 Verbal lexemes

Esperanto has an open list of selective verbs, like *neĝ-i*, 'to snow' (zerovalent), *promen-i*, 'to have a walk' (monovalent), *manĝ-i*, 'to eat' (bivalent), and *lu-i*, 'to rent' (trivalent, the renter being the Experiencer and the syntactic subject). Of course, any lexeme can be verbified, so it will have its own valence value and actant structure, i.e., where to put actants in the adtree.

The verb *esti*

The verb *esti* is a calque frome the Latin *esse*, 'to be'. As seen in section 2.9.2, this verb has a special value in Indo-European languages, as it is bivalent, but its second argument, introduced by Minus, is not an accusative, but always a nominative. This is an exception: some rules are written only to cover this special case. The second argument can be stative (*mi estas homo*, 'I am a human being'), circumstantial (*mi estas frue*, 'I am early'), or adjunctive (*mi estas alta*, 'I am tall').[39] Finally, if the second argument is adjunctive, the verb *esti* can be omitted (see next section).

5.5.6 Adjunctive lexemes

Adjunctive lexemes pertain to precise semantic domains, as for instance colours, *flav-a*, 'yellow', *ruĝ-a*, 'red', or qualities *bel-a*, '{nice|handsome}', *alt-a*, 'high', *bon-a*, 'good'. For lexemes belonging to this semantic domain, the antonymiser *mal-* is exceptionally productive: *mal-alt-a*, 'low', *mal-bon-a*, 'bad', etc.. In the dictionary, adjunctive lexemes are monovalent unless specified otherwise.

- (132a-eo.) Liza estas bela.
- (132b-eo.) Liza belas.
- (132ab-en.) Liza is nice.

Note that the direct verbification of an adjective lexeme, as in example 132b, builds a slightly different construal, i.e., there is a more active flavour of Liza's niceness in example 132b compared to example 132a.[40] In fact, the two adtrees are different.

5.6 Affixes

To be rigorous, affixes should be put as a subsection of the lexicon, as in adtrees they are rightly put on leaves, not on hooks, as adpositions. However, as the Esperanto affix system is quite rich, I decided to put them in an apart section for readability.

Transferees are morphemes that do not change the grammar character of their governor and they are part of the lexicon. Esperanto has a lot of affixes which are highly productive and powerful, because there is almost no allomorphy. Most Esperanto affixes are either prefixes or suffixes.

Some affixes can be applied only to a specific grammar character, e.g., there are static transferees, that can be applied only to stative governors. However, most transferees can be applied regardless of the grammar character: they are called generic affixes. In practice, the categorisation of Esperanto is as follows:

- generic prefixes (λ-prefixes);
- generic suffixes (λ-suffixes);
- verbal prefixes (i-prefixes);
- verbal suffixes (i-suffixes);
- stative prefixes (o-prefixes);
- stative suffixes (o-suffixes);

Table 5.25: The generic prefixes (λ-prefix class)

λ-prefix	**Governor**	**Basic modulus**	**Additional moduli**
mal-	λ	+[ANTONYM]	
ne-	λ	+[LACK]	
...			

In the following, prefixes and suffixes will be presented along their classification. For brevity, I will omit details which are not relevant for the goal of this dissertation part: presenting conventions and rules of the Esperanto adgram.

5.6.1 Generic prefixes

Table 5.25 shows some generic Esperanto prefixes. The prefix *mal-* is an antonomyser, and it can be applied in principle to any lexeme: adjectival (e.g., *bon-a*, 'good', *mal-bon-a*, 'bad'), verbal (e.g., *sukces-i*, 'to succeed', *mal-sukces-i*, 'to fail'), stative (e.g., *ord-o*, 'order', *mal-ord-o*, 'disorder'), circumstantial (e.g., *tro*, 'too much', *mal-tro*, 'not enough'). If applied to adpositions, e.g., *mal-antaŭ*, it is considered an extreme lexicalisation (see section 5.5.3). I have said 'in principle' because in practice it is applied if it makes sense. In other words, there is an open list of morpheme where this prefix can be applied, that cannot be captured by the formal model, as it is not described in terms of grammar characterisation. Furthermore, in some cases the antonymiser is a modulus inside the lexeme itself: *memor-i*, 'to remember', *forges-i*, 'to forget' (it is equivalent to *mal-memor-i*, which is not practically used), *tag-o*, 'day', *nokt-o*, 'night'. In adgrams, these lexical antonomy relations are *not* rendered explicitely, as they are not represented morphologically, and in fact these relations are highly debated among cognitive linguists [Taylor, 2002].

As a stand-alone morpheme, *ne* is clearly an adverb, expressing negation. Its homonymic signature *ne-* espresses lacking. Let us see the series with the lexical base *util-a*, 'use-ful': *mal-util-a*, 'harm-ful', *ne-util-a*, 'useless'.

5.6.2 Generic suffixes

Table 5.26 shows some of the Esperanto verbal prefixes. The suffix *-aĉ-*, officialised in 1909, means 'of bad quality, not adequate' and it is calqued from Yiddish: *dom-aĉ-o*, 'a hovel', *far-aĉ-i*, 'to do badly', *mol-aĉ-a*, 'flabbly, flaccid, limp'. The suffix *-eg-* is an augmentative, and it can be applied

Table 5.26: The generic suffixes (λ-suffix class)

λ-suffix	Governor	Basic modulus	Additional moduli
-aĉ-	λ	+[QUALITY,BAD]	
-eg-	λ	+[AUGMENT]	
-et-	λ	−[AUGMENT]	
-em-	I	+[-WARD,ABSTRACT]	
-em-	O	+[-WARD,ABSTRACT]	+[LIKE]
-et-	O,+[FAMILY]	−[AUGMENT]	+[DEAR]
-ul-	λ	+[ANIMATE,MASCULINE]	+[HABIT]
-il-	λ, −[O,−[INSTR.]]	−[ANIMATE,INSTR.]	
-ad-	λ	+[CONTINUOUS]	+[HABIT]
-ad-	−[O,+[IL]]	+[CONTINUOUS]	+[ACTION]
...			

to governors of any type: *dom-eg-o*, 'a big house', *varm-eg-a*, 'very warm', *kri-eg-i*, 'to shout very loudly'. The form *dank-eg-o-n!* is an idiomatic form meaning 'many thanks'. The suffix *-et-* is complementary: *dom-et-o*, 'a small house', *pluv-et-i*, 'to drizzle'. With family names, it is also used to express endearment: *patr-et-o*, 'dad', *edz-et-o*, 'my dear husband'. There are a few lexicalised forms where it also means endearment: *knab-in-et-o* or *et-ul-in-o*, 'a lovely little girl'.

The suffix *-em-* expresses an inclination towards the moduli brought by the governor. It is mostly attached to verbal lexemes and often it takes part to a suffix chain with the animative suffix *-ul-*: *drink-em-ul-o*, 'drunkard', *mensog-em-ul-o*, 'liar', *decid-em-ul-o*, 'one who is inclined to take decisions'. Attached to stative lexemes, it brings the additional modulus of liking: *muzik-em-a*, 'that loves music', *ver-em-a*, 'that loves the truth'.

The suffix *-il-* indicates instruments and tools, and it is the antonym of *-ul-* concerning animateness. The tools can be of any dimensions and use, as the suffix can be applied to verbal and stative lexemes. Some examples with verbal lexemes: *hak-il-o*, 'a hook' (see again example 122a), *ŝlos-il-o*, 'a key (the tool for closing)', *komput-il-o*, 'a computer', *print-il-o*, 'a printer'. Some examples with stative lexemes: *buter-il-o*, 'something for spreading butter', *fotokopi-il-o*, 'a photocopier'.

There is a special class of inanimate selective nouns which already have the modulus [INSTRUMENT], e.g.: *bros-o*, 'a brush', *martel-o*, 'a hammer', *revolver-o*, 'a revolver', *telefon-o*, 'a telephone'. They are clearly selective nouns because, when verbified, they take the meaning of 'using that tool'. The suffix *-il-* cannot be applied to this special class. Interestingly, the list

Table 5.27: The verbal prefixes (I-prefix class)

I-prefix	Governor	Basic modulus	Additional moduli
ek-	I,+[ACTION]	+[START,NOW]	±[CONTINUOUS]
dis-	I,+[ACTION]	+[DIRECTION,MANY], +[GENERIC]	
...			

Table 5.28: The verbal suffixes (I-suffix class)

I-suffix	Governor	Basic modulus	Additional moduli
-ebl-		+[-IBLE]	
-end-		+[MUST,PASSIVE]	
...			

is somehow limited: two calques of *computer* where proposed in the 1970s, *komputor-o* (analogously to *transistoro*) and *komput-il-o*. Today *komput-il-o* has definitely won.

The suffix *-ad-* and the predix *ek-* are complementary. The basic modulus of *-ad-* is +[CONTINUOUS]. Normally, it is used with verbal lexemes: *uz-ad-o*, '(regular) use', *vizit-ad-o*, 'a visit (the act of usual or regular visiting)'. It is also used with intrumental statives, specifying the action: *martel-ad-o*, 'hammering', *telefon-ad-o*, 'a phone call', *komput-il-ad-o*, 'a computation made with a computer'. If used with other lexemes, it brings the additional modulus +[USUAL]: *mal-fru-ad-o*, 'the act of being usually late'.

5.6.3 Verbal prefixes

Table 5.27 shows the Esperanto verbal prefixes. The prefix *dis-* means 'in generic, many directions'. It can be applied only to verbal lexemes having the modulus [ACTION]: *send-i*, 'to send', *dis-send-i*, 'to send in a mailing list', *hak-i*, 'to hack', *dis-hak-i*, 'to hack in many pieces'.

5.6.4 Verbal suffixes

Table 5.28 shows some Esperanto verbal suffixes. The suffix *-ebl-* is calqued from the Latin suffix *-ibilis*, also present in English, and it means 'that can be done': *manĝ-ebl-a*, 'edible', *leg-ebl-e*, 'that can be read'. It can be used

Table 5.29: The valence suffixes (I-valence-suffix class)

Affix	Governor	Basic modulus	Additional moduli
-ig-	λ,[VALENCE= 0–2]	$x >$I,[VALENCE= $+1$]	$S = \{N_a \vert N_x\}$
-iĝ-	λ,[VALENCE= 2]	$x >$I,[VALENCE= 1]	$S = N_p$
-iĝ-	λ,[VALENCE= 1]	$x >$I,[TRANSFORM]	$S = N_P$

as a lexeme: *ebl-ec-o*, 'possibility', *ne-ebl-a*, 'impossible'. The form *kompren-ebl-e* is an idiomatic expression meaning 'of course, for sure', and it should be listed apart as a single item. The suffix *-end-* is the last affix to be officialised (1953), and it is calqued from Latin: *leg-end-a*, 'that should be read', *pag-end-a*, 'that should be paid'. Note that a group head by the preposition *de* which is dependent from an A-group having inside *-end-* always indicates the Agent, while a dependent group head by the preposition *al* always indicates the Experiencer, e.g., *pagenda de mi al vi*, 'to be paid by me to you'.

5.6.5 Special verbal suffixes

There are two special classes of verbal suffixes, that are treated apartly: valence transferers and participles.

The valence transferers

There is a class of special suffixes with only two elements. Their function is to modify the value of valence and therefore the actant structure in the current phrase. They modify the grammar character, so they should be presented rightly in the adposition section; however, as they are suffixes, please let me present them here, just after ordinary verbal suffixes. Table 5.29 shows the Esperanto valence affixes, one of the most powerful class of affixes of this QNL. First of all, valence suffixes are inline verbifying transferers of the grammar character of the lexeme (or the atom, in case of compounds): this changes the whole set of affixes which can be applied to the lexeme (see section 5.5.2 for further details).

The suffix *-ig-* is calqued by German but it has been generalised. The subject of a verb with *-ig-* is always the Agent or the Instrument, regardless of the original valence: *la nub-o pluv-ig-as*, 'the cloud has caused the raining' (zerovalent verb), *Li sid-ig-as la hund-o-n*, 'he had put the dog seated' (monovalent verb), *la knab-ino aĉet-ig-as donac-o-n al la patr-o*, 'the little girl has made his father buy a gift'. If present, the second valence is the Patient (e.g., *hund-*) and the third valence the Experiencer (e.g., *patr-*). This suffix is never applied to trivalent verbs, such as **donigi*; instead, an analystic strategy is followed (example 133).

- (*133eo.) *N'_a(Alfredo) donigas ??(Bean) doni N_p(libron) N_e(al Karlo).
- (133eo.) N'_a(Alfredo) igas N'_p(Bean) N_a(doni) N_p(libron) N_e(al Karlo).
- (133en.) Al forces Bea to give a book to Charles.

The suffix *-iĝ-* is complementary, and it is of ergative use [Gledhill, 1998, 29]: if the lexeme is originally monovalent, there is no change in valence, but a transformation: *la krem-o glaci-as*, 'the cream is ice(d)', *la nub-o glaci-iĝ-as*, 'the cloud has become a lump of ice', *ŝi pal-iĝ-is*, 'she turned pale'. If it is applied to bivalent verbs, the subject becomes the Patient, while the Agent remains unexpressed, as shown in examples 134,135. Of course, there is no use with zerovalent verb.[41]

- (134eo.) N_p(La lecion-o) $V^{2-1=1}$(komenc-iĝ-as).
- (134en.) The lesson is starting.
- (135eo.) N_a(La profesor-o) V^2(komenc-as) N_p(la lecion-o-n).
- (135en.) The professor is starting the lesson.

The modification in the actant values can be predicted in advance, because the verb valences are known in advance by dictionary. This is particulary important in the case of augmentation (suffix *-ig-*). In example 136, the increased value is N_e because the first valence of the lexeme *komenc-* is N_a while the second one is N_p. Vice versa, in examples 137–138 the third value is N_p because the actants of the lexeme *telefon-* are respectively N_a and N_e – the Instrumental is an extra actant, as explained in section 2.7.1.

- (136eo.) N_a(La profesor-o) V^{2+1}(komenc-ig-as) N_p(la lecion-o-n) N_e(al la tutoro).
- (136en.) The professor makes the tutor starting the lesson.
- (137eo.) N_a(Paŭlo) V^2(telefonas) N_e(al Liza).
- (137en.) Paul telephones Liza.
- (138eo.) N_a(Karlo) $V^{2+1=3}$(telefonigas) N_p(Paulon) N_e(al Liza).
- (138en.) Charles make Paul to phone Liza.

Both suffixes are used together with prepositions in form of prefix, as special verbal paradigms: *sen-vest-ig-i*, 'to take off the cloths', *sen-vest-iĝ-i*, 'to take one's cloths off'; *sen-fort-ig-i*, 'to take off the strenght'; *sen-fort-iĝ-i*, 'to weaken'; *al-bord-ig-i*, 'to take on board', *al-bord-iĝ-i*, 'to take oneself on

Table 5.30: The structure of Esperanto participles

Morpheme	Gov.	Diathesis	Basic modulus
-ant-(-a-j-n)	I	+[ACTIVE]	+[CONTINUOUS]
-int-(-a-j-n)	I	+[ACTIVE]	−[CONTINUOUS]
-ont-(-a-j-n)	I	+[ACTIVE]	+[START,NEXT]
-at-(-a-j-n)	I,[VALENCE= 2]	−[ACTIVE]	+[CONTINUOUS]
-it-(-a-j-n)	I,[VALENCE= 2]	−[ACTIVE]	−[CONTINUOUS]
-ot-(-a-j-n)	I,[VALENCE= 2]	−[ACTIVE]	+[START,NEXT]

Table 5.31: Esperanto circumstantial participles

Morpheme	Gov.	Diathesis	Additional moduli
-ant-e	I	+[ACTIVE]	$S = t$
-int-e	I	+[ACTIVE]	$S = t$
-ont-e	I	+[ACTIVE]	$S = t$
-at-e	I,[VALENCE= 2]	−[ACTIVE]	$S = t$
-it-e	I,[VALENCE= 2]	−[ACTIVE]	$S = t$
-ot-e	I,[VALENCE= 2]	−[ACTIVE]	$S = t$

board'; *al-supr-ig-i*, 'to take someone on the top', *al-supr-iĝ-i*, 'to take oneself on the top' ; *en-dom-ig-i*, 'to take into the house', *en-dorm-iĝ-i*, 'to fall oneself asleep'; *en-tomb-ig-i*, 'to bury (into a tomb), to inter', *en-tomb-iĝ-i*, 'to enter into a tomb, or to bury oneself'.

The participles

A second class of suffixes involved in verb formation are participles. In Esperanto, participles are used mainly to indicate the aspect. Esperanto follows the Slavic model in this respect, which is more rich and symmetric compared to Western Indo-European languages.[42] Table 5.30 shows the basic structure of participles, which are a special case of suffixes. Note the perfect symmetry between the vowels of the participles and the verbal finals (Table 5.8). Their use depends on the grammar character they are applied to. They are never used with compounds. The structure seen before is not changed if they are part of an adjectival group, both if applied to a stative group or to *esti*. In contrast, their use is more rich if they are used as circumstantials or statives, i.e., the basic moduli are retained in any case. Circumstantial participles correspond to gerund in English, *gerundio*

Table 5.32: Esperanto stative participles

Morpheme	Gov.	Diathesis	Additional moduli
-ant-o(-j-n)	I	+[ACTIVE]	+[HUMAN],
-int-o(-j-n)	I	+[ACTIVE]	+[HUMAN],
-ont-o(-j-n)	I	+[ACTIVE]	+[HUMAN],
-at-o(-j-n)	I,[VALENCE= 2]	−[ACTIVE]	+[HUMAN],
-it-o(-j-n)	I,[VALENCE= 2]	−[ACTIVE]	+[HUMAN],
-ot-o(-j-n)	I,[VALENCE= 2]	−[ACTIVE]	+[HUMAN],

in Italian, and sometimes as absolute ablative in Latin. This means, that such a group introduces a new phrase which is dependent from another. There is no use of *-n* with circumstantial participles.

- (139a-eo.) Manĝante pomon, Liza legas.
- (139a-en.) While eating an apple, Liza is reading.
- (139b-eo.) Manĝinte pomon, Liza legas.
- (139b-en.) After eating an apple, Liza starts reading.
- (139c-eo.) Manĝonte pomon, Liza legas.
- (139c-en.) Just before eating an apple, Liza reads.

In example 139, the subject of the dependent phrase is a trace for Liza: this is an information brought by the circumstantial use of the participle. If stative, the participle has the additional modulus of 'a human being doing something' if active, and 'a human being being done something' if passive.

- (140a-eo.) Paŭlo estas la amato de Liza.
- (140a-en.) Paul is Liza's lover.
- (140b-eo.) Liza estas la amatino de Paŭlo.
- (140b-en.) Liza is Paul's lover.

Note that participles block the use of suffixes like *-ul-*, save in lexicalised cases like *kon-at-ul-o*, 'acquaintance', and few others. If inanimation is needed, the suffix *-aĵ-* is added: *leg-at-aĵ-o*, 'something being read', although mostly *leg-aĵ-o* is enough, i.e., without the participle.

Table 5.33: Gender and endearment affixes

Morpheme	Gov.	Basic modulus	Additional moduli
ge-+-j	O,+[MASCULINE]	+[BISEX,BOTH] −[MASCULINE]	
ge-	I	+[BISEX,BOTH] −[MASCULINE]	
-in-	O,+[MASCULINE],	+[FEMININE] −[MASCULINE]	
ge-+-j	O,+[MASCULINE]	+[BISEX,BOTH]	
bo-	O,+[HUMAN]	+[-IN-LAW]	
...			

For joking, it is possible to found *-unt-* participles: e.g., *stud-unt-o*, 'someone who should study but actually he or she is not', a lazy student. However, this use is not part of the standard register of Esperanto. Let me conclude with a side note. Wennergren [2005, 398] correctly notes that *loĝ-at-a*, 'inhabited', is possible, even if in traditional grammar this is considered "intransitive". In fact, in adgrams, the correct basic form of the verb is *loĝi en*, because *en* is a false adposition, being a marker of a Minus relation in a *bivalent* verb. In other words, the formal model is fit for these otherwise puzzling cases.

5.6.6 Stative affixes

Stative affixes can be furtherly divided in affixes applied to animate lexemes and affixes applied to inanimate lexemes.

Gender and endearment

Table 5.33 shows gender and endearment affixes. Animate affixes, as ordinary lexemes, can be marked with the feminine suffix or by the bisexual prefix if they are masculine. Neutral affixes or lexemes – such as animals – cannot be modified by the feminine suffix *-in-*: **hom-in-o* for 'a man', instead of *vir-in-o*; **tabl-in-o*, '*she-table', **meger-in-o*, '*hag'. Some calques of common European proper names show an *-a* as the feminine final: *Paŭl-o*, 'Paul', *Paŭl-a* or *Paŭl-in-o*, 'Paula' or 'Pauline', *Liz-a*, 'Liza', *Ana*, 'Anna', *Barbara*, 'Barbara'. All these are standard lexemes with the modulus [FEMININE] in.

The circumfix *ge-+-j* (the only one in Esperanto) is used with masculine nouns: *ge-sinjoro-j*, 'Ladies and Gentlemen', *ge-frat-oj*, 'brothers and

Table 5.34: The Esperanto animate stative prefixes

Affix	Gender	Governor	Basic modulus	Additional moduli
ĉef-	neutral	O	−[HUMAN],+[HEAD]	
...				

sisters', *ge-patr-oj*, 'the parents'. The homonymous prefix *ge-* is used sometimes with verbal lexemes: *ge-lernej-o*, 'a school for boys and girl', *ge-dorm-i*, 'to have sex heterosexually in bed' (informal).

To indicate endearment within proper and family names, Esperanto has a couple of Slavic suffixes *-ĉj-* and *-nj-* – as in the common Slavic feminine names Tatiana/Tanja, Sofia/Sonja. These two suffixes are inflective, unlike any other Esperanto suffix – hence, they are considered lexicalisations. For masculine nouns, the suffix is *-ĉj-*: *Johan-o*, 'Johnathan', *Jo-ĉj-o*, 'John', *patr-o*, 'father', *paĉj-o*, 'dad'. To express endearment, the suffix *-et-* is also used.

Stative animate prefixes

Table 5.34 shows Esperanto animate stative prefixes.[43] The prefix *ĉef-* is similar in meaning to *-estr-*, but it can been applied to any stative lexeme: *ĉef-redaktor-o*, 'the editor-in-chief', *ĉef-strat-o*, 'the main street'. Note that *-estr-* always indicates a human being, while *ĉef-* not: *urb-estr-o*, 'a mayor', *ĉef-urb-o*, 'the capital city'.

Animate stative suffixes

Table 5.35 shows Esperanto animate stative affixes.

The suffix *-an-* is neutral and it is used to sign membership, within groups: *estr-ar-an-o*, 'a member of the leading group', *famili-an-o*, 'a member of the family', *kongres-an-o*, 'a member of the congress'. Second, it can used with places, that are metaphorically sets of members: *Belg-uj-an-o*, 'someone that lives in Belgium', *Pariz-an-o*, 'a Parisian'. Third, it is used for adepts of religions, philosophies and any doctrine: *Krist-an-o*, 'a Christian', *Islam-an-o*, 'a Muslim'.

The suffix *-estr-* indicates the leader or the head of the place indicated by the governor: *bank-estr-o*, 'a director of a bank', *imperi-estr-o*, 'an emperor', *lern-ej-estr-o*, 'a head teacher'.

The suffix *-id-* is a calque from Greek, and it means [DESCENDANT]: *ŝaf-id-o*, 'a lamb', *arb-id-o*, 'a young tree', *reĝ-id-o*, 'the king's descendant', *Israel-id-o*, 'an Israelite'. A lexicalised form is *Latin-id-a lingv-o*, 'a Romance language'. The suffix *-in-* is explained in section 5.6.6.

Table 5.35: The Esperanto animate stative suffixes

Affix	Gender	Governor	Basic modulus	Additional moduli
-an-	neutral	O,+[SET]	+[MEMBER]	
-an-	neutral	O,+[PLACE]	+[MEMBER]	+[INHABITANT]
-an-	neutral	O,+[DOCTRINE]	+[MEMBER]	+[ADEPT]
-estr-	neutral	O,+[PLACE]	+[HUMAN,] HEAD	
-id-	neutral	O,+[ANIMATE,] LIVE	+[DESCENDANT]	
-in-	mascu.	O,+[ANIMATE,] LIVE	+[FEMININE]	
...				

Inanimate stative suffixes

Table 5.36 shows the Esperanto inanimate stative affixes. The couple of suffix *-aĵ-* and *-ec-* are often put in contrast as antonyms, but their behaviour shows a relation more complex than plain antonymy. The suffix *-aĵ-* is basically a concretive. Applied to adjuncts, it is the complement of the suffix *-ec-*, which is always abstractive, both in quality and in state: *util-aĵ-o*, 'something useful', *util-ec-o*, 'utility', *nud-aĵ-o*, 'something bare or naked', *nud-ec-o*, 'nakedness'. In contrast, if applied to verbal lexemes, it signs a single product of the action, according to the basic modulus of the verbal lexeme: *far-aĵ-o*, 'something done', *desegn-aĵ-o*, 'a picture', *trink-aĵ-o*, 'something to drink'. If applied to something edible, the suffix *-aĵ-* indicates a dish: *suker-aĵ-o*, 'a sweet delicacy'. The suffix *-ec-* applied to stative lexemes shows quality: *hom-ec-o*, 'humaneness', *ŝton-ec-o*, 'the quality being hard as a stone', *viv-ec-o*, 'aliveness'. Finally, sometimes these two affixes are lexicalised: *dolĉ-aĵ-o* is 'a sweet', not a generic 'sweet thing', *konstru-aĵ-o* is 'a building', not generically 'a thing that was built', *nov-aĵ-o* is 'the news', not 'something new'; analogously *bel-ec-o* is 'beauty', not 'something which is nice'. These entries should be listed apart.

The suffix *-ej-* brings the modulus [PLACE], and hence it cannot be easily applied to lexemes already that modulus: *?Franc-uj-ej-o*, '?a place of France', *?urb-ej-o*, '?a place of the city', *?hospital-ej-o*, '?a place for being an hospital' (in those cases, an average Esperantophone would prefer a word compound strategy: *Franc-uj-lok-o*, perhaps the 'Ilê de France', *urb-o-lok-o*, 'a particular place of the city', *hospital-o* or *mal-san-ul-ej-o*, a calque from the German *-e Krankenhaus*, 'a hospital'). A common use is with verbal lex-

Table 5.36: The Esperanto inanimate stative affixes

Affix	Governor	Basic modulus	Additional moduli
-aĵ-	A	−[ABSTRACT]	+[ENTITY]
-aĵ-	I	−[ABSTRACT]	+[ACTION,PRODUCT]
-aĵ-	O,+[EDIBLE]	−[ABSTRACT]	+[DISH]
-ec-	A	+[ABSTRACT]	+[ENTITY]
-ec-	O	+[ABSTRACT]	+[QUALITY]
-ej-	I	+[PLACE]	
-ej-	O,+[OBJECT]	+[PLACE]	+[SHOP]
-ej-	O,+[OBJECT]	+[PLACE]	+[STORE]
-ej-	O,+[PLANT,CULTIV.]	+[PLACE]	+[PRODUCE]
-ej-	O,+[OBJECT,NATURE]	+[PLACE]	+[FULL]
-ej-	O,+[ANIMATE]	+[PLACE,WORK]	
-er-	O,−[ANIMATE]	+[ELEMENT]	
-ar-	O,±[ANIMATE]	+[SET]	
...			

emes, where some forms are lexicalised: *lern-ej-o*, 'a school', not only 'a place where to learn'; *preĝ-ej-o*, 'a Church', not only 'a place where to pray'. [44] With objects, it indicates places where to store or to sell those things. Note that the suffix *-ej-* is deliberately generic: if precision is needed, Esperantophones calques a strategy from a source language. For example: *aŭt-ej-o* can be either a *garaĝ-o*, 'garage', or an *aŭt-o-park-o*; *cigar-ej-o* can be either a shop where they sell cigars, or a room where to store them. With cultivable plants, it indicates the place of cultivation: *tabak-ej-o*, 'a tobacco plantation', *riz-ej-o*, 'a rice-field'. With other natural objects, it indicates a place where that object is abundant: *herb-ej-o*, 'a meadow', *sabl-ej-o*, 'a sandbox', *glaci-ej-o*, 'a glacier'.

The suffix *-er-* indicates the elements of a set composed by inanimate objects, or a natural element seen as a set: *mon-er-o*, 'a coin (literally, a piece of money)', *sabl-er-o*, 'a speck of dust', *neĝ-er-o*, 'a snow flake', *pluv-er-o*, 'raindrop'. The suffix *-ar-* is its complementary: *arb-ar-o*, 'the wood or the forest', *membr-ar-o*, 'the members (of a group)'. However, most of its compounds are lexicalised: *hom-ar-o*, 'humankind', which is more than 'a set of human beings', *bird-ar-o*, 'the birds (all the birds in the world)', *gazet-ar-o*, 'the press', which is more than 'a set of newspapers', *ŝtup-ar-o*, 'a ladder', which is more than 'a set of bars'; *vort-ar-o*, 'dictionary', which is more than 'a set of words', *vagon-ar-o*, 'convoy, train', which is a more than 'a set of wagons'. In case of ambiguities, Esperantohones use word compounds,

Table 5.37: Pronouns in Esperanto

Person	NUMBER	GENDER	O	A
i	singular	λ	mi(-n)	mi(-a-j-n)
ii	singular	λ	vi(-n)	vi(-a-j-n)
iii	singular	masculine	li(-n)	li(-a-j-n)
iii	singular	feminine	ŝi(-n)	ŝi(-a-j-n)
iii	singular	neutral	ĝi(-n)	ĝi(-a-j-n)
i	plural	λ	ni(-n)	ni(-a-j-n)
ii	plural	λ	vi(-n)	vi(-a-j-n)
iii	plural	λ	ili(-n)	ili(-a-j-n)
iii	λ	λ	oni(-n)	oni(-a-j-n)
iii	$t = S_{iii}$	λ	si(-n)	si(-a-j-n)

with stative lexemes like *element-*, *pec-*, 'piece', *part-*.

5.7 Pronouns

Pronouns are placeholders for stative lexemes, most often for actants. In Esperanto pronouns are mostly borrowed directly for English, except for *oni*, which is a calque from French (but also present in German, i.e., *man*), and *si*, which is borrowed from Latin. *Si* is a placeholder of the syntactic subject S of the current phrase, if it is a third person (singular or plural). It can never be used itself as subject.

- (*141a-eo.) *Paŭlo kaj sia amatino dancas.
- (141b-eo.) Paŭlo kaj lia amatino dancas.
- (141b-en.) Paul and his own lover are dancing.
- (141c-eo.) La dancado de Paŭlo kaj sia amatino mirigas.
- (141c-en.) The dancing of Paul and his own lover is amazing.
- (141d-eo.) La dancado de Paŭlo kaj lia amatino mirigas.
- (141d-en.) The dancing of Paul and his lover is amazing.

In example 141d, the construal depicts a scene of dancing of Paul with someone's lover, say, Charles: a totally different situation compared to the other variations of the example!

Table 5.38: The Esperanto invariable determiners

Morpheme	**Basic modulus**	**Additional moduli**
□	−[DEFINITE]	
la	+[DEFINITE]	
unu	+[ONE,ONLY]	±[DEFINITE]
unuj	−[SINGULAR]	−[DEFINITE]
ambaŭ	+[DUAL]	+[DEFINITE]

5.8 Determiners

Under a cognitive point of view, determiners are important bridges between the linguistic and the extralinguistic world. In fact, they often introduce anaphoras and cataphoras. However, adgrams treat them only as purely linguistic objects, for the purposes of this dissertation.

For implementation, Esperanto determiners should be divided between invariable and variable determiners: the last ones should agree in gender and number with the other linguistic objects belonging to the same group.

5.8.1 The invariable determiners

Table 5.38 shows the invariable determiners in Esperanto. Articles are determiners with an apart status in most Indo-European NLs, and Esperanto is no exception. As in Slavic NLs, there is no indefinite article, but there is a unique definite article *la*, borrowed from English. In fact, it is invariable, i.e., it does not change along gender and number, unlike for example Italian, although its use is different from English and perhaps from any existing NL. In particular, Jansen [2007, 274] notes: 'with regard to SVO, I investigated and proved the influence of the definite/indefinite factor on the probability of inversion of the constituents.'

The numeral *unu* is special, and it can be used as a determiner. It is semidefinite, because it is vehicle of two presumptions: the speaker presumes that the hearer does not know the actant to which *unu* is attached. This nuance is not implemented in adgrams by the moment. If pluralised in *unuj*, the meaning completely changes, because even the speaker is not aware of the object of the discourse, and that is why I put it as a different morpheme (for the interested reader, **unun* is ungrammatical). Finally, the form *unu-el-a(-j-n)* is emerged, with the meaning of 'one of a set'. This is a regular adjective, so it should not be treated apartly.

Table 5.39: The Esperanto variable determiners

Subgroup	Pluraliser	Accusative	Basic modulus
ProForma-u(jn)	⊤	⊤	+[DEFINITE]
ProForma-a(jn)	⊤	⊤	+[QUALITY]
ProForma-es	⊥	⊥	+[POSSESS]

5.8.2 The variable determiners

Table 5.39 shows the variable determiners in Esperanto, which are extracted from the Table 5.15. Note that the saxon genitive does not need to follow agreement.

5.8.3 Numerals

Numbers always have an apart place in NLs, and Esperanto is no exception. Being a planned language, it shows a sofisticated and regular system to treat numerals and numbers in general. As in most parts of the grammar, I base myself on Wennergren [2005]: the treatment of time, i.e., dates and hours, are here omitted, as they are not interesting in order to clarify the adpositional grammar model.

From zero to 999,999

There are 13 **basic numbers** in Esperanto that form an apart class:

- *nul* = 0
- *unu* = 1
- *du* = 2
- *tri* = 3
- *kvar* = 4
- *kvin* = 5
- *ses* = 6
- *sep* = 7
- *ok* = 8
- *naŭ* = 9
- *dek* = 10^1

Table 5.40: The numeral suffixes

Suffix	Basic modulus
-obl-	+[-IPLE]
-on-	+[FRACTION]
-op-	+[TUPLE]

- *cent* = 10^2
- *mil* = 10^3

Numeral lexemes

There is a special class of stative lexemes which is used to express big numbers, the **numeral lexemes**:

- *milion-* = 10^6
- *miliard-* = 10^9
- *bilion-* = 10^{12}
- *trilion-* = 10^{18}
- *kvadrilion-* = 10^{24}
- *kvintilion-* = 10^{30}

Bigger numbers are possible, but they are out of the standard language register so they are omitted.

Numeral specifications

Numeral specifications can be formed either by stative basic numbers or by numeral lexemes and they are always introduced by the preposition *da* without the accusative marker: e.g., *miliardoj da formikoj*, 'billions of ants'. This subgroup can be also formed by basic numbers, used as stative lexemes: e.g., *centoj da birdoj*, 'hundreds of birds'. Of course, if it makes sense, any stative lexeme can be used in this way, but not within a numeral group: e.g., *malmulte da tempo*, 'so little time'.

Numeral suffixes

Table 5.40 shows the numeral suffixes. The numeral suffixes can be applied to any numeral lexeme, even if they are mostly found applied to small

Table 5.41: The numeral determiners

Suffix	Basic modulus	Additional moduli
preskaŭ	+[APPROX]	NumAtom(∅), NumAtom(-obl-), NumAtom(-on-),
ĉirkaŭ	+[APPROX]	NumAtom(∅), NumAtom(-obl-), NumAtom(-on-),
plimalpli	+[APPROX]	NumAtom(∅), NumAtom(-obl-), NumAtom(-on-),
pli	+[MORE]	NumAtom(∅), NumAtom(-obl-), NumAtom(-on-),
malpli	−[MORE]	NumAtom(∅), NumAtom(-obl-), NumAtom(-on-),
po	+[PO]	NumAtom(∅)

numbers, rarely to hundreds and thousands and above. [45] The suffix *-on-* introduce a fraction: *du-on-a*, 'half', *ducent milonoj*, '200/1000'; in contrast, the suffix *-obl-* introduces a multiple: *du-obl-o*, 'the double', *cent-obl-e*, 'a hundred times'. Finally, the suffix *-op-* has the same meaning of the English suffix *-ple*: *n-opo*, 'n-tuple', *du-opl-e*, 'as a couple'.

Unlike the *-on-*, the suffix *-op-* can also be used with other stative lexeme, if it makes sense: e.g., *mult-op-e* is Zamenhofian, *Rigardu, kiel multope jam aliĝis al ni la Britoj*, 'see, how many Britons already came to us'. Today's Esperanto would consider *mult-e* enough. However, this use is so limited that those forms can considered lexicalisations.

Numeral determiners

When used as numeral determiners, *preskaŭ* and *ĉirkaŭ* have the same meaning of the lexicalisation *pli-mal-pli*, *pli* meaning 'more' and *mal-pli* meaning 'less'. The pattern *(mal)pli... ol...*, '{more|less}... than... is', is used with ordinary lexemes, so it is parsed appropriately.

The next subsection is dedicated to explain the very special morpheme *po*.

The word of the Devil: *po*

Cherpillod [1996] offers a convincing comparative analysis of one of the most tricky part of the Esperanto grammar, even if he put it erroneously in the category of prepositions. The word *po* is borrowed from Russian and it is present in most Slavic languages, but a similar strategy is used by Japanese too. Let me give extract examples.

- (142-eo.) Mi donis al la infanoj po 3 bonbonojn.
- (142-en.) I gave 3 sweets to the children each.
- (142-it.) Diedi 3 caramelle a ciascun bambino.
- (143-pl.) Dostali po 3 jabłkow.
- (143-hr.) Dobiše po 3 jabuke.
- (143-en.) They received 3 apples each.
- (144-jp.) Bîru-o 3 zucu yatta.
- (144-en.) (Someone) gave 3 glasses of bier each.

Adgrams shows clearly that *po* is a *determiner of numerals*, because it can never put in a phrase along with other determiners: it is completely irrelevant if the group is nominative or accusative, one of the most debated point, as referred by Cherpillod [1996]. Table 5.17 shows the use of determiners of numerals in various Esperanto and in English. Table 5.18 shows the use of determiners of numerals in Italian and in Polish. Finally table 5.19 shows the same example in Japanese and Croatian (note that the adtrees in Croatian and Polish are identical). It is quite clear that Western languages, such as English or Italian, put this special morpheme as a modifier of the Experiencer (N_e), while Slavic languages or Japanese put it as a modifier of the number of the Patient (N_p), which is cognitively clear. Esperanto follows the Slavic example.

It is theoretically possible to combine *po* with the numeral suffixes, even if it seems a bit too odd to an esperantophone. Finally, as a prefix, *po-* is also used, even if it is controversial: e.g., *poiome*, approximately 'step by step', is found 10 times in the whole *Tekstaro de Esperanto*, made of 4,266,767 words (5 December 2008).

5.9 A corpus-based example

Let me conclude this chapter with a rich sentence extracted from the web edition in Esperanto of *Le Monde Diplomatique*:

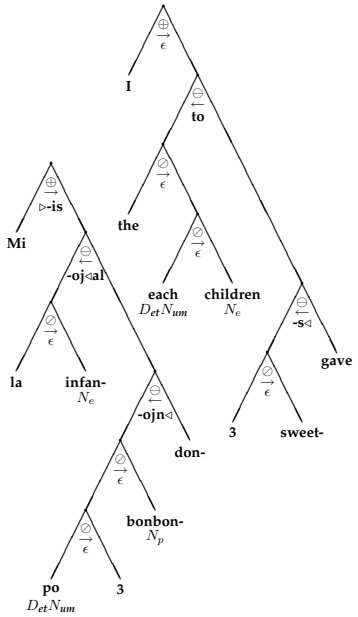


Figure 5.17: Determiners of numerals in Esperanto and in English.

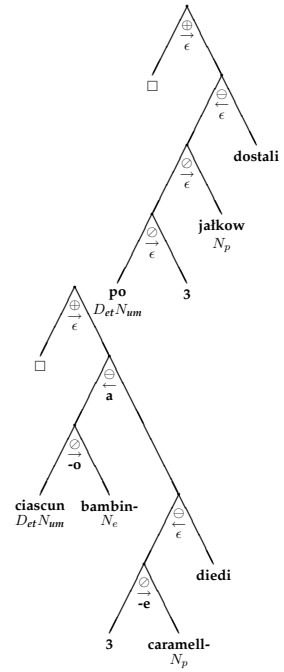


Figure 5.18: Determiners of numerals in Italian and Polish.

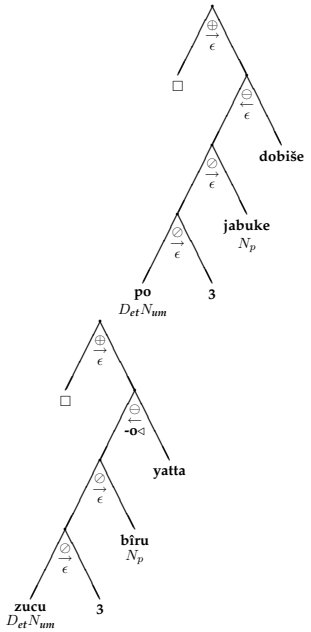


Figure 5.19: Determiners of numerals in Japanese and Croatian.

> Sekve al la franca kaj nederlanda referendumoj de 2005, la traktato pri funkciado de la Eŭropa Unio (TFEU) forlasas la konstitucian leksikon kiu karakterizis la eŭropan konstitucian traktaton kaj kiu havis la celon "eternigi" la liberalajn principojn.[46]

This is a rough translation in English: 'After the French and Dutch referenda of 2005, the Treaty on the Functioning of the European Union (TFEU) let fall the constitutional lexicon that characterised the constitutional treaty and that had the goal of making the liberal principles "eternal"'. This sentence is quite complex, as it is made of three phrases, i.e., a main phrase and two relative clauses. For readability, I split this sentence in two adtrees (see Figures in the following pages: double junctives and correlatives in bold, also the left member, which is omitted in the original text):

- (145main-eo.) Sekve al la franca **kaj** nederlanda referendumoj de 2005, la traktato pri funkciado de la Eŭropa Unio (TFEU) forlasas la konstitucian leksikon //
- (145rel-eo.) // **kiu** karakterizis la eŭropan konstitucian traktaton **kaj kiu** havis la celon "eternigi" la liberalajn principojn.

Figure 5.20 shows the main phrase; in particular, note the external circumstantial opened by *Sekve al* – here exploded in its components, but probably to be considered a lexicalised adposition pattern. In contrast, I considered *Eŭropa Unio* as a lexicalisation, and rendered as *EU*: this is to show that the designer of an adpositional grammar system should take some choices, but always at a linguistic level, i.e., what to consider lexicalisation and hence to be put in the dictionary. Furthermore, please note the double group *franca kaj nederlanda* nested are appended to *referendumoj* in the last position. Figure 5.21 shows the clauses appended at the main phrase thanks to the double junctive *kaj*. They are both opened by the $\mathcal{C}$-correlative *kiu*, which acts as the hook of the syntactic subject S of the respective clauses. The presence of *kaj* is important, because it pushes down the first clause: otherwise the second clause would be attached to *traktat-* instead of *leksik-*.

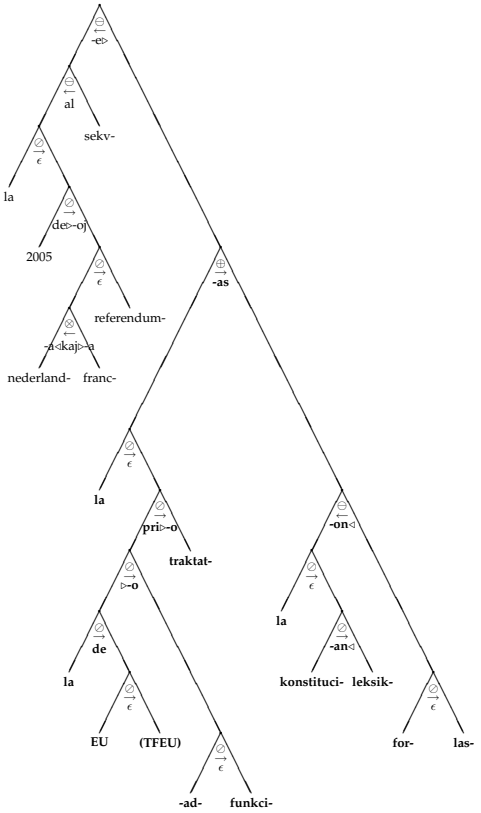


Figure 5.20: The adtree of *Sekve al...* (145, main phrase).

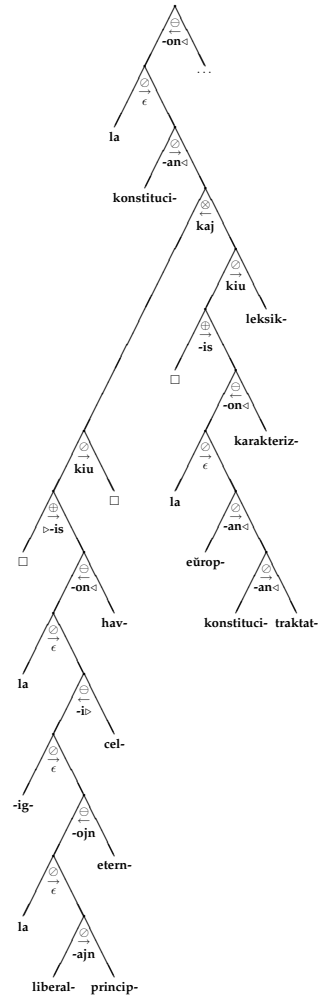


Figure 5.21: The adtree of *Sekve al...* (145, clauses).

Notes

[1] At Zamenhof's time Belarussian was not distinct from Russian; nonetheless, the influence of Belarussian is clear in certain phonological traits. For this relation, see the voice "Esperanto Phonology" of Wikipedia, page version ID: 241992246. Last change: 30 September 2008, CET 11:59. Retrieved 5 November 2008.

[2] There are more subtile and precise classifications of borrowing than the one I have offered here, notably the classic ones by Eugen Haugen and Uriel Weinrich. I limit myself to this very sharp dinstinction as I do not need anything more precise for the purposes of this dissertation.

[3] However, in this case the German-like construal would probably be considered the best form among esperantophones with a mastery level (C2). This is a matter of style. Adgrams are not developed enough to give account of style matters, by the moment.

[4] For the Esperanto grammar, I use mainly Wennergren [2005], the complete reference monolingual grammar, and Gledhill [2001] and Gledhill [1998], the only corpus-based grammar, for control. For specific collocational aspects, my references are Jansen [2008] and Jansen [2007]. Moreover, I use Willkommen [2007], a reference grammar of Esperanto written in German, the Italian edition of the so-called Zagreb method, Imbert et al. [1985], the first international Esperanto course based on a computational analysis of the everyday language corpus, and Migliorini [1995], a classic handbook for Italians, where the grammar is presented in a traditional way. These three references I use for control. For dictionaries, I use mainly Duc-Goninaz [2002], the standard reference monolingual dictionary, although highly criticised. I also use Benson [1995] – the most used English-Esperanto dictionary – and Minnaja [1996] and Broccatelli [1991], the most complete Italian-Esperanto dictionaries. I use English and Italian as they are the source and target languages in the first scenario of the translation game. Finally, I use Krause [1999] and Krause [2007], the most complete German-Esperanto dictionaries, for control.

[5] In particular, I use a corpus-based analysis of Zamenhof's language use [Nomura, 1989], the collected works [Dietterle, 1983] as well as his classic linguistic answers [Itô, 1990]. Furthermore, I use Grosjean-Maupin et al. [1956], a classic monolingual dictionary.

[6] If a morpheme presents different conducts, that morpheme will be counted twice or more, as a special homonymic case.

[7] Esperanto shows some traits that are definitely not Indo-European, as the lexical use of preposition: *per-ant-o* is a present participle, whose lexical head is the preposition *per*. In English, it could be a monstrous word like **through-er*, 'go-between, mediator'. In adgrams, such very special and rare case should be treated as duplications: i.e., there will be an entry in the dictionary for the adposition, and a separate entry for the lexical use.

[8] In writing his grammar of Yiddish, Zamenhof had chosen a Latin-derived alphabet which eventually becomes the alphabet of Esperanto, with very few changes [Holzhaus, 1982]. It is possible to write Esperanto without the diacritics, and in actual practice there are at least two ways to do it. For the purposes of this dissertation, Esperanto has only the official alphabet.

[9] This is the original citation: "cette particularité fournit un moyen mnémotechnique commode, l'espéranto étant plus ou moins une moyenne des différentes langues européennes" Tesnière [1959, 64]. It is hardly surprising that Esperanto syntax may be well represented in terms of dependency grammars.

[10] One can argue that the Esperanto lexeme *formal* has the pseudosuffix *-al*, exactly as in English [Wiener, 1951]. It is certainly true that the Romance languages brought into the Esperanto lexicon a highly inflective pseudoaffix derivation – for instance, the word *neglekt-i*, 'to neglect', can be analysed by back-formation (*retroderivado*) as *ne-glekt-i*, so to obtain a new verbal lexeme, *?glekt-i*, as an antonym of *neglekti* ('to care?'). Nonetheless, revitalising this "linguistic fact", as tried mainly by the Czech writer Karolo Piĉ, did not have any success in the actual use of Esperantophones [Haupenthal, 1987]. Piĉ can be considered a modernist like James Joyce for English. In fact, both exploited the extreme linguistic

tendencies of their literary languages (*La Litomišla tombejo*, *Finnegan's wake*) – see also Sutton [2008, 466–471] about Piĉ.

[11] I will solve this 'almost' in section 5.5.2 about word compounding.

[12] The English indefinite article *a* is rendered with a zero adposition in Esperanto, as in Slavic languages; see section 5.8 for details on articles.

[13] Contra Wennergren [2005] and Willkommen [2007], following Imbert et al. [1985] and Migliorini [1995], I decide to use traditional grammar terms like 'nominative', 'accusative', as this strategy does not cut Esperanto from its source languages, performing the so-called 'propædeutic effect' in language learning, identified in the pioneer article by Eaton [1927] and proved in varius research projects (for a survey, see Pinto [2002] in Di Stefano and Renzo [2002]).

[14] However, *esti*, 'to be', is tricky in any case, but for different reasons, that I explain later.

[15] It is noteworthy that Zamenhof considered *-aŭ* a true adverb final: "[it] does not belong to the root, but it is only a facultative final, so it is omittable, as 'o' in substantives" [Itô, 1990, 5B]. However, this possibility nevere entered the standard language register of Esperanto.

[16] The modulus [DIR.] in *laŭ* and *kontraŭ* means 'direction'.

[17] I thank Iván Bujdosó for giving me the source of the following table [Szerdahelyi, 1989].

[18] For instance, morpheme clusters as *which types of* of example 32d are considered as an idiomatic expression so to express a correlation. The table is modelled on Esperanto: the table for other NLs can be different, but only increasing the number of formas, never the pro formas. As explained in chapter 7, the implementation can be extended if needed.

[19] It is attested some use of circumstantial correlative verbified by the appropriate transfer: e.g. *tiel-as*, 'things are so'; *kiom-as* 'how much?'. This is a case of lexicalisation: the correlative is put as it is in the lexicon, having lost its grammar potentialities. In the adtree, it will be simply put in the appropriate leaf.

[20] It can also used for exclamations [Wennergren, 2005, 224]. This fact does not prejudge the line of reasoning I am presenting.

[21] A very common error, also made by Esperantologists like Migliorini, is to consider the forma *-u* as belonging to the modulus ±[ANIMATE]. It is not so, as this example in very common, plain Esperanto shows clearly.

[22] It is far easier to consider the forma *-e-n* completely grammaticalised, instead of analyse it as +[PLACE,MOVE]. However, the implementation permit to have circumstantial variable correlatives as well, if needed.

[23] Pretending that Esperanto is an "Asian language" – whatever this would really mean, usually meaning NLs like Chinese or Japanese – because of its use of affixes is simply a nonsense – e.g., Piron [1981]. Not only the lexicon is cleary 99% Indo-European – yes, words like *alĝebro* and *algoritmo* are originally arabic, but they are calqued through the mediation of Indo-European languages! – but also its structure is clearly Indo-European, as I am showing at any step of this chapter. In fact, how would be possible for Zamenhof to plan an 'Asian QNL', if its source languages are either Indo-European or Semitic? The true reason of this big lie is political: some Esperantists pretend that Esperanto becomes the world language (*mondolingvo*), and they want to avoid the ideological attack of Eurocentrism (as a side note, not Zamenhof, who postulated in 1915 the United States of Europe, but this is another story). Recently the Japanese Esperantologists Yamasaki [2000] and Tida [2007] in Gotoo et al. [2007] offered additional arguments to my thesis, in this respect.

[24] Further details are given through the implementation in the following chapters.

[25] For the over-formal Esperantologist, in Esperanto there are three cases: nominative, accusative, and genitive, which is fossil, i.e., no longer productive. There is a tentative to revitalise this fossil case in words like *?mies*, 'my', or *?alies*, 'belonging to another person'. These proposals are superflous, because there are already *mia(jn)* and the regular compound *ali-ies*. For details over this very special topic, see the discussion in Wennergren [2005, §13.3].

[26] Last but not least, the definition of 'clause' can be extended to stative groups when governed by a preposition. This will become clear in the formal model.

[27] This is particulary clear with *be-cause*, which is a grammaticalisation having the word *cause* inside. The model is spurious in this specific point.

[28] In the early version of adgrams the property of symmetry seen by Brøndal seemed to be promising: in the end, it was cut off by the model, as junctives show the property of associativity, not symmetry, at a formal level. So, symmetry should be considered a semantic proper and a guideline to put rightly adpositions when building the adspaces of NLs. In a sense, semantic symmetry is carved into Times.

[29] Of course, this list can be expanded whenever the formal model is applied to other language realities.

[30] At the following web addresses:
http://www.akademio-de-esperanto.org/verkoj/baza_radikaro_oficiala.html
and http://wikisource.org/wiki/Baza_Radikaro_Oficiala.
Retrieved 1 December 2008.

[31] One of the numerous reforms of Esperanto proposed in the 20^{th} century was exactly of this character: for a comparison between the lexical approach of Latino sine Flexione, proposed by Giuseppe Peano, and the approach of standard Esperanto, well analysed by Eugen Wüster, see [Gobbo, 2008b].

[32] On a cognitive level, I argue that this last sentence is processed slowly than the previous one, exactly because our mind should calculate the transfers.

[33] The first study of compounding was published in 1910 by René de Saussure and it is now available as an ebook [De Saussure, 2003]: this study, published first in French and then in Esperanto, was clearly approved by Zamenhof himself as a good guideline, especially for writers [Itô, 1990, 4]. René de Saussure was a Swiss mathematician, an early Esperanto supporter, and the friend of the more famous Ferdinand de Saussure. A presentation of his life and works is Künzli [2001] in Fiedler and Liu [2001]. Other sources of analysis is the corpus-based grammar of Esperanto by Gledhill [1998, 31-37] and the study of Cherpillod [2003].

[34] A direct consequence is that I consider compounds like **Lizmarodomo*, '?Liza's house at the seeside', as ungrammatical. It may be that such monstrous compounds do exist in Esperanto, but they are definitely out of the standard register of the language, the one which I am describing and implementing.

[35] However, it is easy to add a rule for circumstantial compounding, if ever needed, as the formalism proposed here is robust enough to do so, if needed.

[36] It seems that the so-called 'adjectival' affixes, to use the imprecise terminology of some Esperanto grammarians, are inline transferers of the lexeme they are applied to. This is of no use from a computational point of view, as the final grammar character is always defined by the Esperanto final system.

[37] Of course, this is valid if the compound is not lexicalised. At the limit, it can be delexicalised, e.g., *akv-o-fal-ad-o* is 'a continous fall of water', while 'a continuos waterfall' is *ad-a akvofal-o*. In other words, in this compound the atom is *fal-* not *akvofal-*.

[38] It is obvious the influence of the source language substrata: *fervojo* is a calque from the French *chemin de fer*, while *malsanulejo* is a calque from the German *Krankenhaus*. For the Esperantologist: it is obvious that the suffix *-uj-* is built semantically in front of *-ej-*. However, I decided to omit it as its intricacies are always at a semantic level, i.e., on moduli, and this does not give any useful information to the model, in particular because of the concurrent forms expressing the relation between inhabitants and countries. Adgrams are agnostic, so they permit at a limit *Afgan-istan-o : Afgan-land-o : Afgan-i-o : Afgan-uj-o*. However, I find the arguments of de Kock [2007] in Löwenstein [2007] very convincing: in spite of all discussions, even in this semantic area Esperanto shows a good degree of regularity compared to NLs.

[39] Note that a frequent error made by Esperantophones is the adjunctivation of the circumstantial, e.g., **mi estas frua*. This is a proof that *esti* need an apart cognitive process, rendered in the formal model as a apart set of rules.

[40] Of course, this is a claim: to be proved, it should be tested on linguistic corporas annotated with a lot of pragmatic information. This is definitely out of the scope of this dissertation.

[41] If needed, the Agent can be expressed by the preposition *de* or equivalent.

[42] There is an important computational linguistics tradition which applies the so-called Scott-Montague semantics, which is a generalisation of Kripke semantics. Montague's thesis is that semantics in NLs is isomorphic to the semantics of formal languages [Montague, 2002] in [Portner and Partee, 2002]. The members of his school compute NLs through quantifiers, existentials, for stativation, and give a sharp description of verbs through formulas to calculate tense and aspects. This part is still to be developed in adgrams, if ever. In fact, my claim is that such representations are too heavy, both cognitively and formally, to be useful for machine (and human) translation.

[43] The prefixes *ge-* and *bo-* may also be considered members of this class; however, for readability I put them in section 5.6.6.

[44] There exists *moske-o* for Muslims, *sinagog-o* for Jews, etc.; the word *kirk-o* for 'Church' – from the German *-e Kirche* – was proposed but it is not really used.

[45] For example, there is only one occurrence of *mil-op-o* in the whole *Tekstaro de Esperanto*, more precisely in *Mortula ŝipo*: *kaj poste sidas milopoj da ili*, 'and then there were thousands of them sitting down'. Today's Esperanto would say more easily *miloj* instead of *milopoj*. Another example: *milion-ono*, 'a millionth', is found only twice in two newspaper articles (4,266,767 words, 8 December 2008).

[46] This is the first sentence of the article *Ĝemelaj traktatoj*, 1 December 2007, by Bernard Cassen. The original is in French.

Part III

Implementation

Chapter 6

The parsing of Esperanto

In this chapter the morphemes presented in the previous chapter will be finally collocated in their groups. This permits the implementation of the Esperanto grammar, i.e., the transformation of the set of Esperanto words forming phrases and sentences in input to the output, i.e., the appropriate adtree(s). The tables that show the parsing patterns of Esperanto are referred in the rules of the formal model (see next chapter). Names help the reader to follow this line of reasoning. A special role is done by the fundamental moduli, which permit to know the final fundamental grammar character of each group. Therefore, the parsing will follow the grammar character taxonomy given throughout the whole dissertation.

Two caveats are needed before to proceed. Of course, the caveat of the name-entity problem stated in the Introduction is still valid: e.g., in the sentence *Andreas studas multe* the word 'Andreas' should be tagged as a name-entity, otherwise it would be reckoned erroneously as a verbal group, being the final *-as* a signature of the fundamental verbal modulus. An additional caveat is of linguistic nature. Namely, this topic – collocation in Esperanto – is still highly debated among specialist. I cannot exclude that rare or 'strange' collocations can exist that do not fix into the model presented hereafter. However, I argue that the model itself captures above the 95% of sentences written in contemporary standard Esperanto. In any case, a systematic testing of the formal model with the available linguistic corpora is out of the scope of this dissertation – perhaps a whole extra PhD dissertation would be needed. Nonetheless, every collocation was manually tested, i.e., the examples offered are extracted from the corpora or immediately derived from them. Finally, please note that my English translations are only for the reader's benefit, because these groups are not collocated in the syntagmatic axis: as stated in chapter 4, translation is possible only within co-text, i.e. with morpheme chains in collocations.

The meaning of the finals in italics will be clarified in section 5.3.1 below. The positions (I_0, I_2, ..., $E_{-1}*$, E_2, ..., A_0* etc.) will be referred in section

Table 6.1: The parsing of λ-subgroups

Subgroups	$P_{refix}*$	S_{pec}	{-o-|-a-|-e-|-i-|$\emptyset$}	A_{tom}	$S_{uffix}*$
Position	λ_5*	λ_4	λ_3	λ_2	λ_1*
Mandatory	$*$	$\perp$	$\perp$	$\top$	$*$

5.3. Positions without star (e.g., I_0) can occur only once within each phrase, i.e., they are singletons, while positions with star (e.g., $E_{-1}*$) can occur indefinitely, albeit in language use this recursion is very limited (hardly more than twice). Some positions are mandatory, others not. The notation is as follows:

- $\perp$: the position is not mandatory, and there can be at best one item;
- $*$: the position is not mandatory, and there can be one or more items;
- $\top$: the position is mandatory, and there is always one item at a time.
- 1^+ : the position is mandatory, and there can be one or more items.

There is always at least one mandatory position which give the final grammar character of the group with a fundamental modulus (see section 5.3.1 below) and gives cues about the adtype, according to the conventions presented in Table 3.9.[1]

6.1 The lambda subgroup

Each table will show the subgroups forming the current group. I call **subgroup** something very generic that concurs into the making of the group itself. It can be a morpheme of a particular type, as well as another subgroup, i.e., a particular morphological pattern that takes place recursively, or even a full-fledged group – in this case, I call it a **nested group**.

An apart status is given to the lambda subgroup, which solves the compounding and the admittable lists of affixes. *λ-subgroups are the main place where to find lexemes in the implementation.* This subgroup is based on the grammar character of the atom (section 5.5.2, that can be modified by the valence transferers *-ig-* and *-iĝ-* (see again Table 5.18 if needed).

Furthermore, the λ-subgroup separates what is on the left from what is on the right in the implementation, i.e., subgroups at its left, for instance prepositions, can be parsed in parallel with subgroups at its right, e.g., predicative adjectives. For details about this point, see the third part of this dissertation.

Table 6.2: The parsing of λ-pron

Subgroups	P_{ron}
Position	$P \rightarrow \lambda_2$
Mandatory	$\top$

Table 6.3: The parsing of $\lambda - t_{race}$

Subgroups	t_{race}
Position	λ_2
Mandatory	$\top$

Table 6.1 shows the model of compounding, that will be applied a lot of times in the instanced groups presented below. For each column, the position is explicited along with its properties clarified above. Furthermore, each table presents some examples. The examples proposed are real Esperanto words, i.e., they are found in the on-line monolingual corpora presented in section 4.3.3, and they are never hapax legomena. In other words, they are really part of the actual language-in-use.[2]

6.1.1 Lambdas and placeholders

There are two special cases of lambdas: pronouns and traces. Pronouns and traces are both placeholders, i.e., they are markers of anaphoras and cataphoras (see again section 3.3.3, if needed). In parsing, Esperanto pronouns take a special place P, which corresponds to λ_2, but it is different, as it deactivates various adjunctive positions – see Table 6.2. As shown in section 5.7, Esperanto pronouns agree with gender, name and case. In contrast, they take directly the place of λ_2, and there is no agreement or any other adposition possible (Table 6.3). Section 6.2.3 will clarify pronouns and traces putting them into action.

6.2 Fundamental groups

Fundamental groups are various instances of one or more λ-subgroups nested one into the others. I call them 'fundamental' because it is obvious to classify them along their grammar characters: verbal, circumstantial, stative and adjunctive. The following subsections give account of the inner structure of each instance of the fundamental groups, while their combination is clarified in the next section.

Table 6.4: The parsing of I-groups

Subgroups	P_{refix^*}	S_{pec}	{-o-|-a-|-e-|-i-|$\emptyset$}	A_{tom}	S_{uffix^*}	final
Example		*demand*	-o-	*sign*		-*as*
	kun-			*vojaĝ*		-*u*
				frenez	-*iĝ-*	-*os*
		daŭr	-*i*-	*pov*		-*is*
Position	I_5*	I_4	I_3	I_2	I_1*	I_0
Mandatory	$*$	$\perp$	$\perp$	$\top$	$*$	$\top$

Table 6.5: The parsing of E-groups

Subgroups	P_{refix^*}	S_{pec}	{-o-|-a-|-e-|-i-|$\emptyset$}	A_{tom}	S_{uffix^*}	-*e*	-*n*
Examples	*mal-*	*mult*	-*a*-	*manier*		-*e*	
	sin-			*don*	-*em-*	-*e*	
				Ital	-*uj-*	-*e*	-*n*
Position	E_5	E_4	E_3	E_2	E_1*	E_0	E_{-2}
Mandatory	$*$	$\perp$	$\perp$	$\top$	$*$	$\top$	$\perp$

6.2.1 Verbal groups

Table 6.4 shows the parsing of verbal groups. The verbal group is always without spaces, i.e., composed only by words. It is found easily through the verbal finals ({-*as*‖-*is*‖-*os*‖-*u*‖-*us*}), which are in position I_i, and fix the rightmost limit of the group. Words with two lexemes are called **compounds** and it will be treated in section 5.5.2 below. The example words are: *demandosignas*, 'put a question mark', *kunvojaĝu*, 'come together having a trip', *freneziĝos*, 'will become crazy', *daŭripovis*, 'that could last'.

6.2.2 Circumstantial groups

Table 6.5 shows the parsing of circumstantial groups, which are found by the mandatory morpheme -*e*. If an E-group is with the E_{-2} empty, it is called an **elementary circumstantial group (e-group)**: this instance of E-group can be applied to verbal groups as well as to generic adjunctive groups – see Tables 6.17 and 6.18 below.

Circumstantial groups are always made by words, i.e., all lexemes are attached one to the others. The examples are: *malmultamaniere*, 'in few manners', *Italujen*, 'to(wards) Italy', *sindoneme*, 'generously (literally, who has tendency to give himself').[3] There is no E_{-1} to remember the mandatory

absence of the pluraliser *-j*.

6.2.3 Stative groups

Stative groups are the most compelling in Esperanto, as they can have any adtype and be loaded of any actant value – see again Table 3.9 if needed. Adjunctive groups are indicated as a whole, and analysed appropriately as 1^{st} or 2^{nd} type (see Tables 6.17 and 6.17 in the following). Analogously, for adjective and a-numeral groups, which are instances of adjunctive groups. Note that, in the adgram formalism, an Esperanto adjective group postponed to the stative lexeme is predicative – i.e., it is always Minus – in analogy to what explained previously with example 61 for Italian in section 3.3.4.[4] Sometimes the presence of the accusative is possible, sometimes its *absence* is mandatory: this is defined via the parsing rules – see chapter 7. A stative group is a member of one of the following classes:

- stative determined invariable group (o-det-inv-groups);
- stative determined invariable correlative group (o-det-inv-corr-groups);
- stative determined variable group (o-det-var-groups);
- stative determined pronominal group (o-det-pron-groups);
- stative pronominal group (o-pron-groups);
- stative numeral group (o-num-groups);
- stative infinitive group (i-groups);
- empty determined invariable group (o-emp-det-inv-groups);
- empty determined variable group (o-emp-det-var-groups).

These stative groups can be subsumed under a generic stative group (O-group). As this is the most complex phrasal group in Esperanto, its explanation is postponed.

Table 6.6: The parsing of o-det-inv-groups

Subgroups	P_{rep}	$D_{et}I_{nv}$	$A_{dj}G_{roup} * \oslash$	$P_{reflx}*$	S_{pcc}	{-o-|-a-|-e-|-i-|∅}	A_{tom}	$S_{uffix}*$	-o	-j	-n	$A_{dj}G_{roup} * \ominus$
Examples	*por*	*la*	*belaj*		*kor*	∅	*jar*		-o	-j		*pasintaj*
	kun	*la*					*amik*	*-in-et-*	-o	-j		
Position	O_8	O_7	A_x*	O_5*	O_4	O_3	O_2	O_1*	O_0	O_{-1}	O_{-2}	A_y*
Mandatory	⊥	⊤	*	⊥*	⊥	⊥	⊤	*	⊤	⊥	⊥	*

Table 6.7: The parsing of o-det-inv-corr-groups

Subgroups	P_{rep}	$D_{et}I_{nv}C_{orr}$	$A_{dj}G_{roup} * \oslash$	$P_{reflx}*$	S_{pcc}	{-o-|-a-|-e-|-i-|∅}	A_{tom}	$S_{uffix}*$	-o	-j	-n	$A_{dj}G_{roup} * \ominus$
Examples		*ties*		*ek-*	*dauŝr*	*-i-*	*pov*		-o		-n	
Position	O_8	O_7	A_x*	O_5*	O_4	O_3	O_2	O_1*	O_0	O_{-1}	O_{-2}	A_y*
Mandatory	⊥	⊤	*	*	⊥	⊥	⊤	*	⊤	⊥	⊥	*

Table 6.8: The parsing of o-det-pron-groups

Subgroups	P_{rep}	AP_{ron}	$A_{dj}G_{roup} * \oslash$	$P_{refix}*$	S_{pec}	{-o-|-a-|-e-|-i-|∅}	A_{tom}	$S_{uffix}*$	-o	-j	-n	$A_{dj}G_{roup} * \ominus$
Examples	*de* *pri* *el*	*niaj* *mia* *ilia*	*malĝentila*		*ĉef* *akv*	∅ *-o-*	*kuzen* *redaktor* *fal*	*-in-* *-et-*	*-o* *-o* *-o*	*-j*		*foraj*
Position	O_8	A_w	A_x*	O_5*	O_4	O_3	O_2	O_1*	O_0	O_{-1}	O_{-2}	A_y*
Mandatory	$\perp$	$\top$	$*$	$*$	$\perp$	$\perp$	$\top$	$*$	$\top$	$\perp$	$\perp$	$*$

Table 6.9: The parsing of o-det-var-groups

Subgroups	P_{rep}	$D_{et}V_{ar}$	$A_{dj}G_{roup} \oslash *$	$P_{refix}*$	L_{exeme}	{-o-|-a-|-e-|-i-|∅}	L_{exeme}	$S_{uffix}*$	-o	-j	-n	$A_{dj}G_{roup}\ominus$
Examples	*en* *al*	*tiun*		*mis-*	*or*	∅	*kandel* *traf*	*-ing-* *-eg-*	*-o* *-o*	*-j*	*-n*	*neforgeseblaj*
Position	O_8	O_7	A_x*	O_5*	O_4	O_3	O_2	O_1*	O_0	O_{-1}	O_{-2}	A_y*
Mandatory	$\perp$	$\top$	$*$	$*$	$\perp$	$\perp$	$\top$	$*$	$\top$	$\perp$	$\perp$	$*$

Table 6.10: The parsing of o-num-groups

S.	P_{rep}	$AN_{um}G_{roup}$	$A_{dj}G_{roup} \oslash *$	$P_{refix}*$	L_{exeme}	{-o-|-a-|-e-|-i-|∅}	L_{exeme}	$S_{uffix}*$	-o	-j	-n	$A_{dj}G_{roup}\ominus$
E.	*por*	*mil* *kvarcent tridek ok* *nul*		*mal-*	*petrol*	∅	*dolar* *oportun* *ŝanc-*	*-aĵ-*	*-o* *-o* *-o*	*-j* *-j* *-j*	*-n*	*malpuraj* *tieajn*
P.	O_8	A_z	A_x*	O_5*	O_4	O_3	O_2	O_1*	O_0	O_{-1}	O_{-2}	A_y*
M.	$\perp$	$\top$	$*$	$*$	$\perp$	$\perp$	$\top$	$*$	$\top$	$\top$	$\perp$	$*$

Table 6.11: The parsing of stative pronominal groups (o-pron-groups)

Subgroups	P_{rep}	P_{ron}	-n	$AN_{um}G_{roup}$
Examples	*pri*	*vi* *ni* *ili*	*-n*	*du* *okdek unu mil tricent sepdek kvar bilionoj da homoj*
Position	O_8	P	O_{-2}	A_z
Mandatory	$\perp$	$\top$	$\perp$	$\perp$

Stative determined groups

Tables 6.6, 6.7, 6.8 and 6.9 show the stative determined groups, which are very similar one with the others, save for agreement. I have chosen to keep them separate, so that they are more easily readable for the reader interested in Esperantology; similarly, the rules will be kept symmetrical. However, it is possible to have less patterns, and from a computational point of view the rules would be more efficient: what is lost is the clearness of the model. I decided to have a clear model than an efficient one. The examples are: *por la jaroj pasintaj*, 'for the years that passed'; *kun la belaj koramikinetoj*, 'with the nice little girlfriends'; *ties ekdaŭripovon*, '{his|her|its} beginning capacity to last (accusative)'; *de niaj kuzeninoj foraj*, 'of our cousins which live far away'; *pri mia malĝentila ĉefredaktoro*, 'about my impolite editor-in-chief'; *el ilia akvofaleto*, 'from their small waterfall'; *en tiun orkandelingo*, 'in(to) that golden candlestick'; *al mistrafegoj neforgeseblaj*, 'to big errors which we cannot forget'.

Stative numeral groups

Table 6.10 shows the stative numeral groups. Note that in this group the atom is always pluralised, even with zero (*nul*). Moreover, if there is a pattern with *unu* without a plural, this *unu* is a determiner, and it will parsed accordingly. The examples are: *por mil petroldolaroj malpuraj*, 'for a thousand petroldollars which are dirty'; *kvarcent tridek ok maloportunaĵoj*, '438 bad conditions'; *nul ŝancojn tieajn*, 'no chances thereby (accusative)'.

Stative pronominal groups

Table 6.11 show stative pronominal groups. They can substitute λ-subgroups in various patterns. The examples are: *vin du*, 'you two (accusative)'; *pri ni*, 'about us'; *ili okdek unu mil tricent sepdek kvar bilionoj da homoj*, 'they, 81,374 billions of human beings'.[5]

Stative infinitive groups

Table 6.12 shows i-groups, which are essentially λ-groups with the signature of infinitive *-i* as the tail. The examples are: *pri ĉirkaŭmeti*, 'about putting around'; *krom vaporŝipigi*, 'instead of put in a steamship'; *surtabligadi*, 'to lay on the table again and again'.

Stative empty groups

These groups have as the lexical base the **trace** of stative lexemes, i.e., an anaphoric or cataphoric anchor. For this reason, I call them stative empty groups: paradoxically, they are stative without showing any 'true' stative

Table 6.12: The parsing of i-groups

Subgroups	P_{rep}	P_{refix}	S_{pec}	{-o-|-a-|-e-|-i-|∅}	A_{tom}	S_{uffix}	-*i*
Examples	*pri*	*ĉirkaŭ-*		∅	*met*		-*i*
	krom		*vapor*	∅	*ŝip*	-*ig*-	-*i*
	sur-				*tabl-*	-*ig-ad*-	-*i*
Position	O_8	O_5*	O_4	O_3	O_2	O_1*	O_0
Mandatory	$\bot$	$*$	$\bot$	$\bot$	$\top$	$*$	$\top$

Table 6.13: The parsing of o-emp-det-var-groups

Subgroups	P_{rep}	$D_{et}V_{ar}$	AP_{ron}	$A_{dj}G_{roup} * \oslash$	$AN_{um}G_{roup}$	t_{race}
	por	*tiuj*	*iliaj*	*almozaj*		*t*
		tiaj	*niaj*			*t*
Position	O_8	O_6	A_w	A_x*	A_z	O_0
Mandatory	$\bot$	$\top$	$\top$	$*$	$\bot$	$\top$

lexeme, but only placeholders, such as traces or pronouns. Table 6.13 shows the empty determined variable group, which is analogue to the empty determined invariable group, save for the agreement. Figure 6.1 shows the adtree of example 145, which has a stative determined invariable pattern nested in.

- (145eo.) Ni kunvenas kun la pasintaj koramikinetoj.
- (145en.) We meet with the past little girl-friends.

The adtree is complete, i.e., also abstract, unused subtrees are shown, along with their position, so to make the structure clear. The other stative determined groups are very similar: their adtrees are left to the reader. The examples are: *por tiuj iliaj almozaj*, 'for that begging of theirs'; *tiaj niaj*, 'ours of that quality'; *al tiuj viaj plenkreskaj ĝentilaj*, 'al those your adult polite ones'; *la ŝiaj*, 'hers'. Table 6.13 shows the empty determined variable group. In this case, $D_{et}I_{nv}$ is always the definite article *la*. This is the only case where adjunctive pronouns and the article can be put together within the same group.

The generic stative groups

A generic, stative group (O-group) subsumes all stative groups seen before. Figure 6.2 shows a generic abstract adpositional tree. Under a complemen-

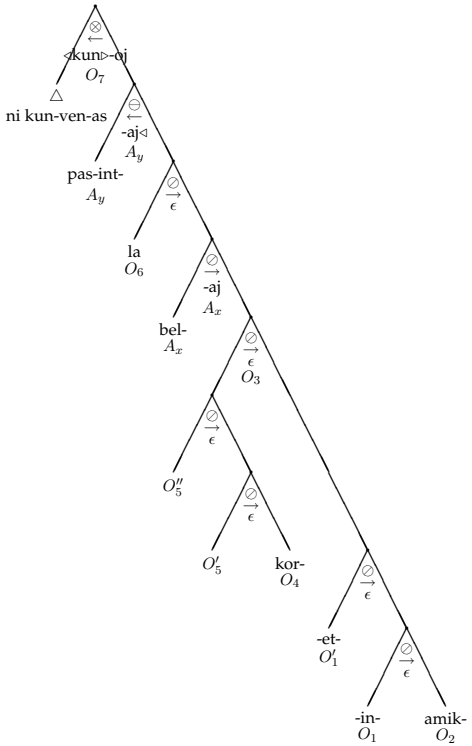


Figure 6.1: The adtree of stative determined invariable groups (ex. 145).

Table 6.14: The parsing of o-emp-det-inv-groups

Subgroups	P_{rep}	$Det\,Inv$	AP_{ron}	$A_{dj}G_{roup} * \oslash$	$AN_{um}G_{roup}$	t_{race}
	al	*la*	*viaj*	*plenkreskaj ĝentilaj*		*t*
		la	*ŝiaj*			*t*
Position	O_7	O_6	A_w	A_x*	A_z	O_0
Mandatory	$\bot$	$\top$	$\top$	$*$	$\bot$	$\top$

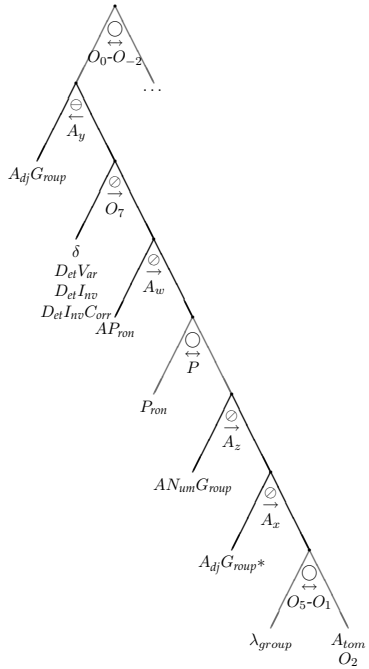


Figure 6.2: The abstract adtree of O-groups.

Table 6.15: The parsing of a-pron-groups

Subgroups	P_{ron}	*-a*	*-j*	*-n*
Example	*mi-* *ni-*	*-a* *-a*	*-j*	*-n*
Position	A_{10}	A_9	A_8	A_7
Mandatory	$\top$	$\top$	$\bot$	$\bot$

tary point of view, this abstract adtree is a generalisation of Figure 6.1. δ is an abstraction to indicate all instances of determiners. Hooks indicates not adpositions but collocational positions, so the most willing reader can check the Tables of stative groups in the adpositional tree: note the lexical base, i.e., the Atom, which is the rightmost participant in the adtree.

6.2.4 Adjunctive groups

Adjunctive groups are the modifiers of stative groups. This means that where they are *always* nested in generic stative groups, as seen before. They are more complex than verbal and circumstantial ones and less complex than stative groups. An adjunctive group is a member of one of the following classes:

- adjunctive pronominal groups (a-pron-groups);
- adjunctive numeral groups (a-num-groups);
- elementary adjectival groups (e-adj-groups);
- adjectival groups (adj-groups);
- generic adjunctive groups (a-groups);
- swapped adjunctive groups (a-swap-groups).

Adjunctive pronominal groups

A particular subset of adjunctives are a-pronouns (see section 5.7). Table 6.15 shows the parsing of adjunctive pronominal groups, which are identified by a couple of mandatory morphemes. There can be at most one pronominal group within an A-group. The example words *miajn* and *nia* mean respectively 'my' – e.g., *vi havas miajn ŝuojn* 'you have my (accusative, plural) shoes' and 'our', e.g., *nia frato*, 'our brother'.

6.2.5 Adjunctive numeral groups

Numerals are identified by a linguistically limited class of morphemes. Numeral groups are singletons in each adjunctive group, exactly as pronominal group. Numbers have a special place in (Q)NLs, as their linguistic planning, i.e., how to name them in speech, was never done with the consultancy of mathematicians, and not rarely *against* them, e.g., the political power of zero was considered highly dangerous by the Church in mediaeval Europe.[6]

As seen in section 5.8.3, Zamenhof has chosen to borrow the model of representing numbers and their modifiers from the East Europe, and planned carefully this part so to have no exception – in contrast, in Italian '16' is *se-dici*, '6+10', while '17' is *dicia-sette*, '10+7', instead of **sette-dici*. Nevertheless, I tried to build the rules so to implement the Esperanto conventions explained, then I realised that the finite-state automaton necessary to implement this part of the grammar is huge and completely useless for three good reasons.[7]

First, whoever will *write* in a numeral group like *plimalpli ducent tridek unu mil kvincent naŭdek unu trilionojn da homoj*, 'more or less 231,591 trillions of human beings', instead of *plimalpli 231.591 trilionojn da homoj*? In the translation game this can be asked to Alice without any problem: if she agrees to play the game and write in Esperanto, for sure she will also write numbers as ciphers, instead of words. In fact, – and this is the second reason – the adtrees of numbers are useless, as there is no explicit adposition to build them, only epsilons (ϵ)! Let me explain through a somehow complex example: *kvarcentope*, 'in 400-tuples', where the lexical base is always void ($\square$), while *kvarcent*, '400', is a compound, and *-op-* is a suffix. Figure 6.3 shows the corresponding adtree. The only relevant information is the suffix *-op-* and the grammar character *-e*: so I decided to retain that information, letting numbers being what they are, i.e., numbers, thanks to their representation in ciphers. The third part of my argument is that the final goal of adgrams is machine translation: it is always possible to write *kvarcent tridekope* as *430-ope*, regardless of the (Q)NL. In this way, numbers are considered as special λ-groups (see section 6.1, if needed) implemented directly in the dictionary. Figure 6.4 shows how the adtree is built. This is the only part of the formal model that implements a part of the encyclopædia, preferring an a priori storage instead of computation, always remember the rule/list fallacy. The additional moduli presented in Table 5.41 are put to avoid that *po* is used if a numeral suffix is already in place.

More scientific uses of numbers, e.g., degrees, are not covered here, as they are not part of the everyday language. In any case, they can be solved following the same approach.

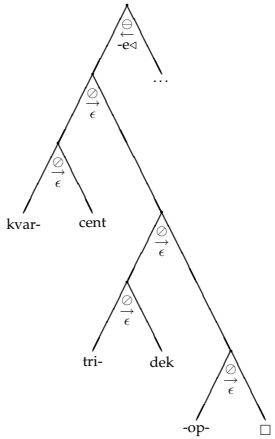


Figure 6.3: The useless adtree of *kvarcent tridekope*.

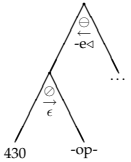


Figure 6.4: The reasonable adtree of *kvarcent tridekope*.

6.2.6 Adjectival groups

The most frequent subset of adjunctives are elementary adjectival groups. Table 6.16 shows the parsing of elementary adjectival groups. Basically, an elementary adjectival group is what expected: a λ-group with the appropriate fundamental, number and case signatures.

As explained immediately below, theoretically there can be an infinite number of them within each adjunctive group. The first example is *longedaŭrajn*, 'that lasts for a long time (accusative, plural)'. This is an intermediate step of grammaticalisation: *longe daŭra* → *longedaŭra* → *longdaŭra* (see section 5.5.2 for this phenomenon, especially important for stative groups). The other examples are: *plimalpli kverelema*, 'more or less badly quarrelsome', *kvar neplaĉaj*, 'four somebody does not like', *pli junan*, 'younger (accusative)'.

Generic adjunctive groups

Tables 6.17 and 6.18 two possible parsings of adjunctive groups in Esperanto. The swapping is rare but Zamenhofian: the example is taken from the novel *Marta*, found in the *Tekstaro de Esperanto*. It is noteworthy that circumstantial groups can act as adjunctive group modifiers, so they should be attached to the appropriate adtree, as shown in the adjective group *foje bongustaj*, 'sometimes tasteful' – Figure 6.5. In conclusion, it is possible to

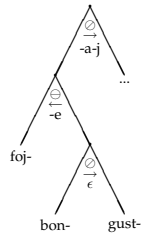


Figure 6.5: The adjective group *foje bongustaj*.

say that, under a cognitive and formal point of view, common linguistic minimal structures seem to be associative. In fact, the only exception is this last swapping, which is commutative.

6.3 The parsing of affixes

In the previous tables, the objects P_{refix} and S_{ufix} were used. In Esperanto, some chains of affixes are allowed, others not. The first criterion

Table 6.16: The parsing of e-adj-groups

Subgroups	P_{refix}*	S_{pec}	{-o-|-a-|-e-|-i-|∅}	A_{tom}	S_{uffix}*-	-a	-j	-n
Examples *plimalpli*				*kverel*	-aĉ-	-a		
kvin milionoj	*ne-*	*mult*		*plaĉ*		-a	*-j*	
pli				*foj*		-a		
				jun		-a		-n
Position	A_5*	A_4	A_3	A_2	A_1*	A_0	A_{-1}	A_{-2}
Mandatory	*	⊥	⊥	⊤	*	⊤	⊥	⊥

Table 6.17: The parsing of A-groups

Subgroups	$AP_{ron}G_{roup}$	$AN_{um}G_{roup}$	e_{group}*	$A_{dj}G_{roup} \oslash *$
Example	*niaj*	*po dek*	*foje*	*plej bongustaj*
Position	A_w	A_z	$E_4 - E_0$*	A_x*
Mandatory	⊥	⊥	⊥	1^+

Table 6.18: The parsing of a-swap-groups

Subgroups	$AN_{um}G_{roup}$	$AP_{ron}G_{roup}$	e_{group}*	$A_{dj}G_{roup} \oslash *$
Example	*du*	*miaj*	*tre*	*pli junaj*
Position	A_z	A_w	$E_4 - E_0$*	A_x*
Mandatory	⊥	⊥	*	1^+

Table 6.19: The parsing of affixes within λ-subgroups (specimen)

p_{ii}	p_i	{O|I}	A_{tom}	s_i	s_{ii}	s_{iii}	s_{iv}	s_v	A_{gr}
bo-	ge-	O	patr						-o(-j-n)
		O	rus			-uj-		-an-	-o(-j-n)
		O	knab		-in-		-et-		-o(-j-n)
		O	ĵurnal-	-ist-	-in-				-o(-j-n)
		O	ŝaf	-id-					-o(-j-n)
	ge-	I	lern-ej-					-estr-	-o(-j-n)
	mal-	O	san	-ul-	-in-	-ej-			-o(-j-n)
	mal-	I	drink	-em-	-ul-				-o
		I	demand				-ar-		-o
		I	leg	-ebl-					-i
		I	vizit	-ad-	-ant-				-i
		I	edz			-in-	-iĝ-		-i

is, again, the grammar character of the affix: e.g., a verbal prefix cannot be applied to a stative lexeme, unless a valence transferer is applied (section 5.6). There is no explicit convention set by Zamenhof about this point, but only his linguistic examples, and after him the style of good Esperanto authors [Sutton, 2008]. Therefore, rules can be extracted inductively basing on language corpora. Sometimes a combination of affixes becomes highly productive, so that the resulting meaning is no more the 'sum' of the morphemes, but a new synthesis – another case of lexicalisation (section 5.5.2).

Table 6.19 shows a specimen of the collocation of affixes. The parsing algorithm treats apartly prefixes and suffixes, after the determination of the grammar character of the Atom (if there is a compound, the grammar character of the Spec is irrelevant). For example, if the grammar character of the Atom is O, the admitted list of affixes is: λ-prefixes; O-prefixes; λ-suffixes; O-suffixes. Analogously if the grammar character of the Atom is I; else, only λ-affixes can be applied. The grammar character of the Atom influences the computation in the λ-subgroup, not the final grammar char-

Table 6.20: The parsing of phrases (sentence level)

Subgroups	$(e_{group}\|O_{group})*$	I_{group}	$(E_{group}\|O_{group})*$
Mandatory	$*$	$\top$	$*$

acter of the group itself, as seen in section 6.1. As a side note, please notice the lexicalisation of *lern-ej-*, 'school', which has acted a permanent transfer I>O.

As a proof, I offer a list of Esperanto words that break these rules: **ge-bo-patr-o*, **san-ul-ant-o*, **knab-ebl-a*, **ge-knab-in-o-j*. Every Esperantophone sees these words as bad-formed, and this is reflected in the formal model. Of course, there are some dependencies between affixes, as in the case of *ge-+-j* (see again Table 5.33). These should be computed with ad hoc rules, which are strictly language-dependent.[8]

The examples in Table 6.19 are: *bogepatroj*, 'parents-in-law', *rusujano*, 'a Russian', *knabineto*, 'a little girl', *ĵurnalistino*, 'a journalist (presswoman)', *ŝafido*, 'a lamb', *gelernejestro*, 'a headmaster (of a school open to both boys and girls)', *malsanulinejo*, 'a hospital only for women', *maldrinkemulo*, 'an alcohol-fobic', *demandaro*, 'a set of question (if lexicalised, a questionairre)', *vizitadanto*, 'a person who is used to be a visitor', *edziniĝi*, 'to marry (literally, to become a wife)'.

6.4 The parsing of phrases and beyond

As stated in section 6.2.4, A-groups are always carved into O-groups. Hence, in order to form a phrase, there are only three generic groups directly involved:

1. the **I-group**, which actives valences and hence the intra-phrasal structure (at the extreme, it would be void: □);
2. at least an **O-group**, filling valences, perhaps enriching the phrasal structure (with zerovalent verbs, it would be void: □);
3. zero or more **E-groups**, which are out of valences – i.e., they always enrich the phrasal structure with additional, external information.

As shown in Table 6.20, the parsing algorithm at this level is now trivial.[9] I call this level the **sentence level**, which is the highest level concerning Esperanto. Above this level there is a **text level**, i.e., the level that transforms a set of sentences in a text. In fact, the rules of the text level are valid regardless of the (Q)NL: a text seems to follow an linear order, otherwise anaphoras and cataphoras simply cannot be solved.

In the next chapter, the rules of the sentence level will be offered, with all its inner layers, which have been sketched herem, as well as the sentence level, which is far simpler indeed.

Notes

[1] To be rigorously, the fundamental moduli are the only ones which are really mandatory, as they can even be used alone as lexemes, as in the often quoted literary example *se mastro as fore*, 'if a master is out' [Cherpillod, 2003]. However, this potential use of the Esperanto grammar is never used practically, at least in the standard register of the language. Nonetheless, the formal model can parse correctly even zero lexemes.

[2] Sometimes they are found only as stative forms, e.g., *demandosigno* ('question mark'), instead of verbal form, i.e., *demandosignas* ('put a question mark').

[3] The prefix *sin-* is in reality the reflexive pronoun in the accusative form *si-n*. However, here it is used as a prefix, so in adgrams it will be considered apart from the reflexive pronoun, analogously to the case of prepositions used as prefixes – see Wennergren [2005, §38.4.3] for details about this extra affixes.

[4] As put clearly by Jansen [2007], this point of Esperanto collocation is still debatable; however, I have take this decision for clearness in the formal model.

[5] 81,374 is the numeric example put by Leibniz to show how to have a universal manner to pronounce numbers in his *lingua generalis* (1678), i.e., *mubodilefa*, or *bodifalemu* [Eco, 1993].

[6] A good survey of the social history of numbers is still Boyer [1968].

[7] In a previous version of this dissertation there was a table for that automaton, which now is left to the over-formal reader.

[8] Under a cognitive point of view, there may be some semantic blocks carved into the lexemes. For example, it is possible to be an ex, i.e., a former spouse, but it is not possible to be an *ex-son or an *ex-parent. Therefore, the English prefix *ex* in the English adpositional grammar should be treated apartly so to be parsed correctly.

[9] As said before, complex phenomena as mirroring or clef sentences are not completely covered, at least in this dissertation. E.g., *Parizen mi veturas.*, 'it is towards Paris, that I am driving by car.', which is a monophrasal sentence formed by [E_{group},I_{group},I_{group}], cannot be parsed at the moment.

Chapter 7

The formal model

In this chapter the formal model is explained in details, so that the virtual Von Neumann's machine can parse texts supposed to be written in standard contemporary Esperanto and the appropriate adtree(s) are generated.[1] Let me recall the two key concepts necessary in order to build an adtree: (a) the **hook**, which is the set of adpositions that heads the current adtree; (b) the **lexical base**, which is the rightmost participant in a given subtree, and it pertains the lexicon. A lexical base can be:

- a lexeme, in simple words (section 6.1);
- an atom, in compound words (section 5.5.2);
- a pronoun, in stative pronominal groups (Table 6.11);
- a trace, in empty groups (section 6.2.3).

The purpose of this chapter is to explain how the adtrees are constructed from the token list derived by the lexical analyser. Since the process that builds adtrees is non-deterministic and is best described by means of derivation rules, I define the parsing algorithm as a set of derivation rules. The next subsection introduces the syntax of rules starting from a standard logical syntax. The logical syntax is extended by adding definitions, in order to model the relevant entities and concepts. Then, the format of rules is presented, explaining their meaning in terms of a non-deterministic computation. Finally, the parsing rules for Esperanto are given, following the guidelines developed in the previous chapters.

These rules form a non-deterministic algorithm since a well-defined computational engine, described along with the formalism, can use them to calculate the adtrees from a token list.

7.1 The virtual machine

The parsing algorithm operates on a virtual machine which can be implemented on a real architecture. Thus, I assume to have a von Neumann's architecture with enough memory.[2] Furthermore, I assume that there is an imperative language available for that machine. A C-like pseudo-language to develop the low-level algorithms is used [Kernighan and Ritchie, 1988].

The distinguishing feature of this parser is its intrinsical non-determinism. As current practice, the non-deterministic behaviour is simulated by means of **backtracking**. [3] When an arbitrary choice has to be taken, one possibility is chosen: if it reveals to be unsuccessful, the machine retreats its choice and inspects another possibility. The backtracking mechanism is well-known and various alternatives to implement it has been studied.[4] Here, the choice is to have a backtrack stack, which is a data structure holding the alternative states of the computation.

The data structure is manipulated by means of two primitives. The backtrack primitive takes one or more parameters. The first one is a label, indicating a point in the code; the other parameters are assignments. The execution of backtrack (label, $v_1 := e_1$, $v_2 := e_2$, ...) pushes onto the backtrack stack the label, the set of all variables in the current state of execution along with their actual values except for v_1, v_2, ... whose values are the result of evaluating e_1, e_2, ..., respectively. Therefore, the entry on the backtrack stack represents an alternative state where the program counter equals the label and the accessible memory equals the stored variables.

The second primitive is **failure**, which takes no parameters. When a failure occurs, i.e., when the failure primitive is executed, the system pops from the top of the backtrack stack a label along with the pairs (v, e) where v is a variable and e is an expression. The execution restarts from the label while the memory is instantiated so to contain exactly the variables stored in the backtrack stack.

Therefore, since the backtracking mechanism is a variant of standard methods, although it is not efficient, it can be implemented on a von Neumann's machine. In other words, a program written using these primitives in an imperative language is computable.

The lexical analysis transforms a sequence of characters in a sequence of **tokens**. Each token is formed by a list of **tags**: each tag identifies an Esperanto modulus as defined in the previous chapters.[5] The lexical analiser is shown in details in the next section, using the described architecture. Hence, the lexical analysis is computable.

The parsing algorithm operates on a virtual machine built on the top of the described architecture. It is similar to Prolog, although a closer resemblance may be found with the so-called logical frameworks.[6]

The virtual machine consists of a complex logical syntax, which allows

to write the parsing rules, plus an inferential engine which operates by means of a satisfiability algorithm that checks whether a rule can be applied to the current state. The details can be found in the section deputed to describe the parsing algorithm.

Since this virtual machine is based on well-known algorithms (namely, satisfiability and matching) and similar systems, e.g., the logical frameworks, have been implemented on real computers, I claim that my virtual machine is computable on the given von Neumann's architecture. Moreover, this claim can be easily proved by implementing the parsing rules in a logical framework: although this has not been done yet, it is evident that the implementation is almost straightforward, regardless of the logical framework used.

7.2 The lexical analyser

The input to the parser is a string of text which is supposed to be written in Esperanto. The goal of the lexical analysis is to convert this string into a list of tokens, representing the whole text whose tokens are, roughly speaking, the morphemes.

The lexical analysis is performed in four subsequent steps:

1. the input string is transformed into a list of words, separating the basic elements;
2. the list of words is scanned to search for composed words;
3. the list of words is cleaned by removing the words with no meaning, like spaces;
4. each word is transformed into a sequence of *tokens*, that is, a set of tags denoting the lexical status of a morpheme in the word.

7.2.1 Step 1: String to list of words

The goal of the first step is to transform the input string into a list of words, where a **word** is understood to be a non-empty string of consecutive characters formed by at least one morpheme between two separators. In adgrams, a word has no special linguistic meaning, it has only a computational meaning, while the linguistic meaning is carried out by the morphosyntax and semantics as explained in the previous chapters.

I assume to have a set, called `WordSeparator`, that contains the characters separating words. This set, containing characters like blanks, commas, periods, etc., is shown in Table 7.1.[7] This set is not rigid, but it is stored in the dictionary. Technically speaking, a word separator appears

Table 7.1: The characters acting as separators

Character	Meaning
.	full stop
	space
?	question mark
!	exclamation mark
,	comma
:	colon
;	semicolon
-	hyphen
/	slash
...	dots
()[]{}	brackets
'	apostrophe

in the dictionary as any other morpheme, but it is marked with the special tag WordSeparator. Thus, the `WordSeparator` set is obtained by querying the dictionary, asking to retrieve all the morphemes tagged with WordSeparator.

A word is defined to be a string of characters, either a separator or the list of characters between two separators. New strings can be constructed by appending characters at the end, via the $\circ$ operation. Similarly, lists can be extended by adding an element to the end, via the $\circ$ operation. The empty string and the empty list are represented by the nil element. No confusion arises since typing allows to decide what is the right interpetation of nil and $\circ$ case by case. The output is a list of words.

The code which parses the string obtaining the list of words is pretty standard. It is shown in Figure 7.1.

7.2.2 Step 2: Alternative interpretation of words

The goal of the second step is to compose sequences of words that may be thought as a single unit. This composition is not compulsory, so it generates an alternative computation track.

Composable sequences of words are represented in the dictionary as a list of regular expressions.

A query to the dictionary is performed by means of the Composable(w) function which returns a list of lists of regular expressions; the result of this function contains the list $r_1, \ldots, r_n$ if w matches the regular expression r_1. A w corresponds to a tag, while e is the regular expression.

An example of a composable sequences of words is the English word '21.7', which is a sequence of three words, '21', '.' and '7'. In order to treat this example, I assume that the following list of regular expressions

```
INPUT( s : string );
i : integer;
w : string;
l : List(string);

w := nil; l := nil;
for ( i := 0; i < len(s); i := i + 1 ) do
    if s[i] ∈ MuteWordSeparator then
        if w ≠ nil then l := l ∘ w;
        l := l ∘ (nil ∘ s[i]); w := nil;
    else w := w ∘ s[i];
enddo
if w ≠ nil then l := l ∘ w;
OUTPUT( l : List(string) );
```

Figure 7.1: Separating the input string in words

is stored in the database: [: *digit*][: *digit* :]$*$ then [: *period* :] and finally [: *digit* :][: *digit* :]$*$.

Matching of regular expressions is defined in the usual way [Wirth, 1986], and I assume that it is implemented via the matchRegexp(w,e) function which matches the e regular expression versus the w word.

The pseudocode deputed to compose words is shown in Figure 7.2.

7.2.3 Step 3: Eliminating mute separators

The list of words obtained so far contains some useless words, which have no meaning.

Table 7.2: The characters acting as mute separators

Character	Meaning
	space
-	hyphen

In fact, some separators are mute and they are recognised because they are members of the set MuteWordSeparator, shown in Table 7.2. Mute separators, e.g., spaces, do not form words. Hence, the goal of this step is to remove mute separators from the list of words. As usual, mute separators are stored in the dictionary database. A morpheme in the dictionary is a mute separator if it is tagged both with WordSeparator and MuteSeparator.

The code deputed to eliminate mute separators is shown in Figure 7.3.


```
        INPUT( l : List(string) );
        ls : List(List(string));
        s, l', l'' : List(string);
        w: string;
        i, j, k, h : integer;
        r : List(List(regexp));
        p : boolean;

→ step2: for (s := nil, i := 0; i < len(l); i := i + 1 ) do
            r := Composable(l[i]);
            if r ≠ nil then
                for (l' := nil, j := 0; j < i; j := j + 1 ) do
                    l'[j] := l[j]; enddo
                for (j := 0; j < len(r); j := j + 1 ) do
                    if i + len(r[j]) < len(l) then
                        for (w := l[i], p := true, k := 1;
                              p ∧ k < len(r[j]); k := k + 1) do
                            if matchRegexp(l[i + k], r[j][k]) then
                                w := w ∘ l[i + k];
                            else p := false;
                        enddo
                        if p then
                            l'' := l'; l''[i] := w;
                            for (h := i + k + 1; h < len(l); h := h + 1 ) do
                                l''[h − k] := l[h]; enddo
                            backtrack(step2, l := l'');
                        endif
                    endif
                enddo
            endif
            s := s ∘ l[i];
        enddo
        OUTPUT( s : List(string) );
```

Figure 7.2: Composing words

```
INPUT( s : List(string) );
i : integer;
l : List(string);

for ( l := nil, i := 0; i < len(s); i := i + 1 ) do
    if s[i] ∉ MuteWordSeparator then l := l ∘ s[i];
enddo
OUTPUT( l : List(string) );
```

Figure 7.3: Removing mute separators

7.2.4 Step 4: words to tokens

The list of words is not enough to parse the text; what is really needed is to classify the single words according to their syntactical nature, in particular their grammar character. This is done looking for them or their components in the dictionary. In fact, a word may be composed by many parts, each one being a morpheme.

The set of information collected for every morpheme is called a token: a token is an entry in the dictionary – see chapter 3 for details. Each token is composed by a set of tags, each one being a modulus. Tags can be written as a single identifier, or as a parametric identifier, like rep(m), where the parameter is a string. The tag rep(m) represents the entry in which the entry is referred with the appropriate grammar character, i.e., how it is written, and it is also used as a key.[8]

The algorithm that transforms words into tokens is shown in Figure 7.4. It operates by scanning every word from left to right, searching for morphemes. When a candidate morpheme is found in the dictionary, inspected via the searchMorpheme function, a list of alternative interpretations of the morpheme is returned. The algorithm takes care of either ignoring these interpretations continuing to expand the morpheme – the first backtrack alternative – or to consider each interpretation – the second backtrack alternative. Each alternative is a token, that is, a set of tags whose textual representation is the morpheme. The list of so-recovered tokens is the output of the lexical analyser.

7.3 The parsing formalism

For the purposes of implementation, a token is a set of **tags**, a tag being the implementation of a modulus. Given a list of tokens $T = t_1, \ldots, t_n$, I say


```
INPUT( s : List(string) );
i, j, k, m : integer;
T, d : List(Token);
f, w, q : string;
b : boolean;

for (T := nil, i := 0; i < len(s); i := i + 1) do
    for (w := s[i], b := true, k := len(w) − 1; k ≥ 0; b := false) do
        for (d := nil, f := nil, j := k; j ≥ 0 ∧ d = nil; j := j − 1) do
  → ext: f := w[j] ∘ f; q := f;
           if (j > 0) then q := "-" ∘ q;
           if b then q := q ∘ "-";
           d := searchMorpheme(q);
           if (d ≠ nil) then
               backtrack(ext, d := nil, j := j − 1);
               m := 0;
      → alt: backtrack(alt, m := m + 1);
               if (m ≥ len(d)) then d := nil;
           endif
       enddo
       if (d ≠ nil) then T := T ∘ d[m];
       else failure();
    enddo
enddo
OUTPUT( T : List(Token) );
```

Figure 7.4: Transforming words into tokens

that $a \leftharpoondown b$, read as "a is strictly on the left of b", and $b \rightharpoonup a$, read as "b is strictly on the right of a", iff there are $h, k \in \{1, \ldots, n\}$ such that $h < k$ and $a = t_h$ and $b = t_k$.[9] Similarly, I write $a \overset{=}{\leftharpoondown} b$ as a shortand for $a \leftharpoondown b \vee a = b$, and $a \overset{=}{\rightharpoonup} b$ as a shortand for $a \rightharpoonup b \vee a = b$.

Furthermore, $\overrightarrow{t} = t_{i+1}$ if $t = t_i$ and $i \leq n$ in the token list $T = t_1, \ldots, t_n$, while $\overrightarrow{t} = \emptyset$ otherwise. Simmetrically, $\overleftarrow{t} = t_{i-1}$ if $t = t_i$ and $i > 1$, while $\overleftarrow{t} = \emptyset$ otherwise. Note that we distinguish the empty token $\emptyset$ from the void token $\square$, denoting an empty leaf, since the last one contains at least an element, namely $\mathsf{Rep}(\epsilon)$, i.e., the empty string ϵ is its textual representation.[10]

The rules are written using formulae based on 'sided' quantifiers, i.e., quantifiers acting on the left or on the right of a specified token. Given a token list $T = t_1, \ldots, t_n$,

- $\forall_{\text{left}(t)} x.P(x)$ is equivalent to $\forall x \in T.x \leftharpoondown t \rightarrow P(x)$.
- $\forall_{\text{right}(t)} x.P(x)$ is equivalent to $\forall x \in T.x \rightharpoonup t \rightarrow P(x)$.
- $\forall_{(s,t)} x.P(x)$ is equivalent to $\forall x \in T.x \leftharpoondown t \wedge x \rightharpoonup s \rightarrow P(x)$.
- $\forall_{(s,t]} x.P(x)$ is equivalent to $\forall x \in T.x \overset{=}{\leftharpoondown} t \wedge x \rightharpoonup s \rightarrow P(x)$.
- $\forall_{[s,t)} x.P(x)$ is equivalent to $\forall x \in T.x \leftharpoondown t \wedge x \overset{=}{\rightharpoonup} s \rightarrow P(x)$.
- $\forall_{[s,t]} x.P(x)$ is equivalent to $\forall x \in T.x \overset{=}{\leftharpoondown} t \wedge x \overset{=}{\rightharpoonup} s \rightarrow P(x)$.
- $\exists_{\text{left}(t)} x.P(x)$ is equivalent to $\exists x \in T.x \leftharpoondown t \wedge P(x)$.
- $\exists_{\text{right}(t)} x.P(x)$ is equivalent to $\exists x \in T.x \rightharpoonup t \wedge P(x)$.
- $\exists_{(s,t)} x.P(x)$ is equivalent to $\exists x \in T.x \leftharpoondown t \wedge x \rightharpoonup s \wedge P(x)$.
- $\exists_{(s,t]} x.P(x)$ is equivalent to $\exists x \in T.x \overset{=}{\leftharpoondown} t \wedge x \rightharpoonup s \wedge P(x)$.
- $\exists_{[s,t)} x.P(x)$ is equivalent to $\exists x \in T.x \leftharpoondown t \wedge x \overset{=}{\rightharpoonup} s \wedge P(x)$.
- $\exists_{[s,t]} x.P(x)$ is equivalent to $\exists x \in T.x \overset{=}{\leftharpoondown} t \wedge x \overset{=}{\rightharpoonup} s \wedge P(x)$.

A **state** is a set of ranges; a **range** is a named pair of tokens $w(r, s)$ both appearing in the token list, such that $r = t_i$, $s = t_j$ and $i \leq j$, and w is an name.

- $k(x, y)$ where k is a constant, x and y are variables, satisfied by any range $w(a, b)$ in the state having $w = k$; moreover, the satisfying assignment poses $x = a$ and $y = b$;
- $k(r, s)$ where k is a constant, r and s are tokens, satisfied by any range $w(a, b)$ having $w = k$, $a = r$ and $b = s$;

- $k(x, s)$ where k is a constant, x is a variable and s is a token, satisfied by any range $w(a, b)$ having $w = k$, $b = s$; moreover, the satisfying assignment poses $x = a$;
- $k(r, y)$ where k is a constant, r is a token and y is a variable, satisfied by any range $w(a, b)$ having $w = k$, $a = r$; moreover, the satisfying assignment poses $y = b$;
- $v(x, y)$ where v is a variable, x and y are variables, satisfied by any range $w(a, b)$ in the state; moreover, the satisfying assignment poses $v = w$, $x = a$ and $y = b$;
- $v(r, s)$ where v is a variable, r and s are tokens, satisfied by any range $w(a, b)$ having $a = r$ and $b = s$; moreover, the satisfying assignment poses $v = w$;
- $v(x, s)$ where v is a variable, x is a variable and s is a token, satisfied by any range $w(a, b)$ having $b = s$; moreover, the satisfying assignment poses $v = w$, $x = a$;
- $v(r, y)$ where v is a variable, r is a token and y is a variable, satisfied by any range $w(a, b)$ having $a = r$; moreover, the satisfying assignment poses $v = w$ and $y = b$.

A statement on the state is a logical formula over ranges interpreted as satisfiable atoms.

The special notation $t \in r(a, b)$, where r, a and b may be variables, means that, given the token list $T = t_1, \ldots, t_n$, if $a = t_j$ and $b = t_k$ under any satisfying assignment on the state, then there is an index i such that $j \leq i \leq k$ and $t = t_i$, or, in human terms, t occurs in the range $r(a, b)$.

A tree is written in the notation defined in the previous chapters, but it is intended here to be encoded as a data structure. In contrast to standard trees, there is the hook, an extra node which lies "above the root", presented since the introduction. The hook identifies the adposition which dominates the tree, or – in more operative terms – it is the tool to attach the current tree on the other tree. The adposition node is denoted through square brackets aside the root of the tree. It is omitted when it is the empty token $\emptyset$.

A statement on a tree is a tree where a node may contain a variable for a subtree. A statement S on a tree is satisfied if the current tree is identical to S except that variables may become instantiated, eventually augmenting the satisfying assignment. By representing trees as terms, the satisfiability of trees correspond to the standard matching problem for terms.

A rule is written as

$$\dfrac{\begin{array}{llll} & \triangleright P_1 & \triangleright P_2 & \\ & \triangleright T_1 & \triangleright T_2 & \\ \triangleright P & \vdots_{\pi_1} & \vdots_{\pi_2} & \cdots \\ \triangleright T & \triangleright P'_1 & \triangleright P'_2 & \\ \triangleright Q & \triangleright T'_1 & \triangleright T'_2 & \end{array}}{\begin{array}{l} \triangleright P' \\ \triangleright T' \end{array}} \text{Rule name}$$

where P is a statement on the current state, T is a statement on the current tree, and Q is a logical formula which may contain variables shared with P and T, while P' is a state update and T' is a tree; moreover, P_i and P'_i are states, while T_i and T'_i are trees. Each P_i and T_i may contain variables shared with P, T and Q and P_j, P'_j, T_j and T'_j for every $j < i$. Intuitively, a **subderivation** π_i may depend only on previous subderivations, i.e., on derivations appearing on its left. Subderivations may be absent, but if they are present, there is a finite and fixed number of them.

The action of a rule is to deduce new values for the current state and the current tree. The rule can be applied if the premises P, Q and T satisfy the current state and the current tree – as explained before – and every subderivation can be completed. Generally, there is more than one way to satisfy the premises P, T and Q; it is obvious that there is a finite number of ways to satisfy them since the current state, tree and token list are finite objects. Each satisfying instance corresponds to a different way to interpret the phrase modelled by the token list. Under a computational point of view, each different instance corresponds to a branch of a non-deterministic computation. Deterministically, all the possibilities are tried until the successful one is found; since the engine supports backtracking it is not too difficult to simulate the non-deterministic behaviour by means of backtracking.

Updates of the current state are written as a set of terms of the form $-r(a, b)$ or $+r(a, b)$, indicating that the range $r(a, b)$ must be deleted from or added to the current state, respectively.

When a rule contains one or more subderivations $\pi_1, \ldots, \pi_n$, the rule can be applied if all subderivations can be carried out successfully. A subderivation π_i is successfully completed if it is possible to derive by means of the parsing rules the state P'_i and a tree satisfying the tree statement T'_i, starting from the initial state P_i and the initial tree T_i.

The initial state of the parsing procedure is the set $\{\text{init}(t_1, t_n)\}$ associated to the token list $T = t_1, \ldots, t_n$ calculated by the lexical analyser and the initial tree is the empty tree.

Rules are applied as far as possible; when no rule can be applied anymore, the current state S and the current tree W are used to decide whether

the process has terminated with a success or with a failure. The parsing algorithm is successful when S is empty, i.e., there is no range left to parse and the W tree does not contain empty leaves.

If the parsing has been successful, the parsing algorithm provides W as its output and then proceeds as if there was a failure; if there has been a failure, the parsing algorithm pops from the backtrack stack an alternative and restarts from there; if there are no alternatives to a failure, the parsing algorithm stops. The result is a sequence of adtrees each one representing a possible interpretation of the input string; if the output is empty, the input string is not in the language.

7.4 Recognising syntactical objects

Every parsing strategy is based on attributing to a group of tokens some syntactical status.

- (76-en). The book costs ten euros.

In example 76, *The book* is expected to be recognised as the syntactic subject, while *costs* should be recognised as the verb and *ten euros* should be labelled as the syntactic object.

In a formal model having a computational meaning, like this one, a compact way to express the basic entities of our grammar is needed, in order to recognise them. More specifically, I need a set of predicates of the form $P(u, v, \ldots)$, i.e., of logical formulae satisfied when the tokens from u to v, included, that forms the syntactical entity I want to determine. In a more human-readable form, these predicates have already been introduced in the previous chapters but, now, we need them as logical formulae to be used to write the parsing rules. The extra parameters are used to extract significant information from the range of tokens.

7.4.1 The λ predicate

The λ-subgroup models the essential block of morphemes that constitutes the base of most groups (see section 6.1).[11]

The whole group is composed by an atom possibly preceded by a specifier, or, simply, it is empty.

$$\begin{aligned} \lambda\text{-group}(u, v) \quad \equiv \quad & \exists a, n.\lambda\text{-atom-group}(u, v, n, a) \vee \\ & \vee (\exists_{[u,v]} w.\exists l.\lambda\text{-spec-group}(u, w, l, a) \wedge \\ & \qquad \wedge \lambda\text{-atom-group}(\overrightarrow{w}, v, n, a)) \vee \\ & \vee u \rightarrowtail v \ . \end{aligned} \tag{7.1}$$

The specifier is composed by l, a fundamental adposition, preceded by a lexeme, the specifier, and a list of prefixes. The adposition may be empty. If present, the adposition must be preceded by an explicit specifier.

$$\begin{aligned}\lambda\text{-spec-group}(u,v,l,a) \quad \equiv \quad & (\exists q.\mathsf{Rep}(l) \in v \wedge \mathsf{Fundamental} \in v \wedge \\ & \wedge\, \mathsf{Atom}(q) \in \overleftarrow{v} \wedge \\ & \wedge\, \lambda\text{-prefix-group}(u, \overleftarrow{\overleftarrow{v}}, a)) \vee \\ & \vee\, (\exists q.l = \emptyset \wedge \mathsf{Atom}(q) \in v \wedge \\ & \wedge\, \lambda\text{-prefix-group}(u, \overleftarrow{v}, a)) \vee \\ & \vee\, (l = \emptyset \wedge \lambda\text{-prefix-group}(u,v,a)) \ .\end{aligned} \tag{7.2}$$

The prefix is a possibly empty sequence of prefix tokens. Prefixes can be generic or, depending on the grammar character of the atom, they may be verbal (I) or stative (O).

$$\begin{aligned}\lambda\text{-prefix-group}(u,v,a) \quad \equiv \quad & u \rightarrowtail v \vee \\ & \vee\, (\lambda\text{-prefix-group}(u, \overleftarrow{v}, a) \wedge \\ & \wedge\, ((a = \text{I-atom} \wedge \mathsf{I\text{-}prefix} \in u) \vee \\ & \vee\, (a = \text{O-atom} \wedge \mathsf{O\text{-}prefix} \in u) \vee \\ & \vee\, \mathsf{G\text{-}prefix} \in u)) \ .\end{aligned} \tag{7.3}$$

The atom is parsed by combining its lexeme with the suffixes. Also, the atom group calculates the valence as the natural valence of the verb plus a modifier induced by the suffixes.

$$\begin{aligned}\lambda\text{-atom-group}(u,v,n+\delta,a) \quad \equiv \quad & \mathsf{Val}(n) \in u \wedge \mathsf{Atom}(a) \in u \wedge \\ & \wedge\, \lambda\text{-suffix-group}(\overrightarrow{u}, v, \delta, n, a) \ .\end{aligned} \tag{7.4}$$

The suffix group is a possibly empty sequence of elements, each one being a suffix. The special valence suffixes *-ig-* and *-iĝ-* modify the valence, as explained in section 5.6.5. Suffixes must agree with the grammar character of the atom, as prefixes do (see section 5.5.2 for a detailed linguistic

description).

$$\begin{aligned}\lambda\text{-suffix-group}(u,v,\delta,n,a) \quad \equiv \quad & (\mathsf{Rep}(\text{ig}) \in v \wedge n \leq 2 \wedge \delta = 1 \wedge \\ & \wedge\, \lambda\text{-ssuffix-group}(\overrightarrow{u}, v)) \vee \\ & \vee\, (\mathsf{Rep}(\text{iĝ}) \in v \wedge n = 2 \wedge \delta = -1 \wedge \\ & \wedge\, \lambda\text{-ssuffix-group}(\overrightarrow{u}, v)) \vee \\ & \vee\, (\mathsf{Rep}(\text{iĝ}) \in v \wedge n = 1 \wedge \delta = 0 \wedge \\ & \wedge\, \lambda\text{-ssuffix-group}(\overrightarrow{u}, v)) \vee \\ & \vee\, (\lambda\text{-suffix-group}(\overrightarrow{u}, v, \delta, n, a) \wedge \\ & \wedge\, ((a = \text{I-atom} \wedge \mathsf{I\text{-}suffix} \in u) \vee \\ & \vee\, (a = \text{O-atom} \wedge \mathsf{O\text{-}suffix} \in u) \vee \\ & \vee\, \mathsf{G\text{-}suffix} \in u)) \vee \\ & \vee\, (u \rightarrowtail v \wedge \delta = 0) \ .\end{aligned} \tag{7.5}$$

If there is a special suffix modifying valence, then the other suffixes should be modified accordingly.

$$\begin{aligned}\lambda\text{-ssuffix-group}(u,v) \quad \equiv \quad & u \rightarrowtail v \vee \\ & \vee\, (\lambda\text{-ssuffix-group}(\overrightarrow{u}, v) \wedge \\ & \wedge\, ((a = \text{I-atom} \wedge \mathsf{I\text{-}suffix} \in u) \vee \\ & \vee\, (a = \text{O-atom} \wedge \mathsf{O\text{-}suffix} \in u) \vee \\ & \vee\, \mathsf{G\text{-}suffix} \in u)) \ .\end{aligned} \tag{7.6}$$

7.4.2 I-predicates

The I-predicate describe verbal groups. A range represents a verb if and only if it satisfies the I-group predicate. If this is the case, that range *is* a I-group. An I-group is a λ-subgroup followed by a signature of the verbal grammar character, which is eventually an adposition. The predicate extracts the valence of the verb, modified by the special suffixes, if present. Thus, the I-group takes two additional parameters: n, the valence, and d the verbal adposition, like *-as* in *pluvas*.

This fairly clear situation is complicated by *esti*, 'to be', that in Esperanto – as in most Indo-European languages – follows different rules than the other bivalent verbs: the second valence cannot be accusative in case (see section 5.5.5).[12] The adjective must agree in number and case with the subject: this agreement is captured via the c argument. Thus, the predicate

recognising a I-group is as follows.

$$\begin{aligned}\text{I-group}(u,v,n,d,c) \quad \equiv \quad & (\mathsf{I} \in v \wedge \mathsf{Rep}(d) \in v \wedge \\ & \wedge \text{I-comp-group}(u, \overleftarrow{v}, n)) \vee \\ & \vee (\exists c'.(c = c' \circ \text{-n} \vee c = c') \wedge \\ & \wedge n = 1 \wedge \mathsf{Rep}(\text{est-}) \in u \wedge \\ & \wedge \mathsf{I} \in \overrightarrow{u} \wedge \mathsf{Rep}(d) \in \overrightarrow{u} \wedge \\ & \wedge (\text{Adj-group}(\overrightarrow{\overrightarrow{u}}, v, \ominus, c') \vee \\ & \quad \vee \text{e-group}(\overrightarrow{\overrightarrow{u}}, v) \vee \\ & \quad \vee \text{nom-stative-group}(\overrightarrow{\overrightarrow{u}}, v))) \ .\end{aligned} \tag{7.7}$$

Except for extracting the valence from the lexeme acting as the root of the verb, the I-comp-group is a λ-group.

$$\begin{aligned}\text{I-comp-group}(u,v,n) \quad \equiv \quad & \exists a.\lambda\text{-atom-group}(u,v,n,a) \vee \\ & \vee (\exists_{[u,v)} w.\exists l.\lambda\text{-spec-group}(u,w,l,a) \wedge \\ & \quad \wedge \lambda\text{-atom-group}(\overrightarrow{w}, v, n)) \ .\end{aligned} \tag{7.8}$$

7.4.3 O-predicates

A stative group is a noun with a precise function within a phrase. Stative groups can be distinguished in **valence arguments** and **extra arguments**. The valence stative groups are:

1. the first valence argument (subject *S*);
2. the second valence argument (object *O*);
3. the third valence argument (dative *D*).[13]

In Esperanto, extra arguments can be stative groups, where the hook can be a preposition. As this case is deeply relevant in order to write predicates and rules, I call these particular stative groups **prepositional clauses**: analogously to correlative clauses, prepositional clauses are pseudophrases acting as adjectives or circumstantials (see section 5.3.9 for $\mathcal{C}$-correlatives).[14]

The most fundamental stative group is an O-group, so I call this set of predicates 'O-predicates'. A stative group is either a simple stative group (O-S-group), or a composition of stative groups. The p parameter represents the **main adposition**. More precisely, it is the one that determines the adtype: in Esperanto, this is mostly the preposition. Under a formal point of view, the main adposition is the one that is used to attach the adtree corresponding to the stative group to the governing adtree. Visually, the main adposition is never 'pushed down' by a left triangle ($\triangleleft$) or a right one

($\triangleright$). The formal model does not need to specify left or right direction at this level; therefore, I introduce the symbol **diamond** ($\Diamond$) in rules, so to make clear if it is possible to have such secondary adpositions, along the main one, or not.[15]

$$\begin{aligned}\text{stative-group}(u,v,p) \quad \equiv \quad & \text{stative1-group}(u,v,p) \vee \\ & \vee (\exists_{(u,v)} x.\exists c.\exists p_1,p_2,c_1,c_2,c. \\ & \quad p = p_1 \Diamond p_2 \Diamond (c_1 \odot_c c_2) \wedge p \neq \bot \wedge \\ & \quad \wedge \mathsf{OPreJunct} \in u \wedge \mathsf{Rep}(c) \in u \wedge \\ & \quad \wedge \text{stative1-group}(\overrightarrow{u}, \overleftarrow{x}, p_1 \Diamond c_1) \wedge \\ & \quad \wedge \mathsf{OPostJunct}(c) \in x \wedge \\ & \quad \wedge \text{stative1-group}(\overrightarrow{x}, v, p_2 \Diamond c_2)) \ .\end{aligned} \tag{7.9}$$

$$\begin{aligned}\text{stative1-group}(u,v,p) \quad \equiv \quad & \text{O-S-group}(u,v,p) \vee \\ & \vee (\exists_{(u,v)} x,y.\exists p_1,p_2,c_1,c_2,c. \\ & \quad p = p_1 \Diamond p_2 \Diamond (c_1 \sharp_c c_2) \wedge \\ & \quad \wedge \text{stative1-group}(u, \overleftarrow{x}, p_1 \Diamond c_1) \wedge \\ & \quad \wedge \text{OJunct}(x,y,c) \wedge \\ & \quad \wedge \text{stative1-group}(\overrightarrow{y}, v, p_2 \Diamond c_2)) \ .\end{aligned} \tag{7.10}$$

Agreement

It is time to introduce the rules for agreement, which can be also used with some A-groups, as explained later. Two operations are needed, as sometimes the pluraliser *-j* and the case marker *-n* need a signature of a fundamental modulus (tag: Fundamental), sometimes not – described in chapters 5,6.

The $\odot$ operation, parametrised by c, is depicted in the following table:

$\odot_c$	$\epsilon/\emptyset$	-j	-n	-jn
$\epsilon/\emptyset$	c	-j	$\bot$	$\bot$
-j	-j	-j	$\bot$	$\bot$
-n	$\bot$	$\bot$	$c \circ$ -n	-jn
-jn	$\bot$	$\bot$	-jn	-jn

The $\sharp$ operation, parametrised by c, is described in the following table:

$\sharp_c$	$\epsilon/\emptyset$	-j	-n	-jn
$\epsilon/\emptyset$	c	-j	c	-j
-j	-j	-j	-j	-j
-n	c	-j	c	-j
-jn	-j	-j	-j	-j

The $\odot$ and the $\sharp$ operations take care of calculating the right agreements among two joined groups and the junction. The interested reader is warmly invited to check how they do their job.

Simple stative groups

A simple stative group is either the composition of fundamental blocks (an O^+-group) or a preposition followed by an O^+-group.

$$\begin{aligned} \text{O-S-group}(u,v,p) \quad \equiv \quad & \text{O}^+\text{-group}(u,v,p) \vee \\ & \vee (\exists q,r.p = q \lozenge r \wedge \\ & \quad \wedge \mathsf{Rep}(q) \in u \wedge \mathsf{Prep} \in u \wedge \\ & \quad \wedge \text{O}^+\text{-group}(\overrightarrow{u},v,r)) \ . \end{aligned} \tag{7.11}$$

A composition of fundamental blocks, the O^+-group, is a fundamental stative group (O-group) or a pair of O^+-groups joined together by an appropriate junction.[16]

$$\begin{aligned} \text{O}^+\text{-group}(u,v,p) \quad \equiv \quad & \text{O-group}(u,v,p) \vee \\ & \vee (\exists_{(u,v)} x,y.\exists p_1,p_2,c. \\ & \quad \wedge p = p_1 \odot_c p_2 \wedge p \neq \bot \wedge \\ & \quad \wedge \text{O}^+\text{-group}(u,\overleftarrow{x},p_1) \wedge \\ & \quad \wedge \text{OJunct}(x,y,c) \wedge \\ & \quad \wedge \text{O}^+\text{-group}(\overrightarrow{y},v,p_2)) \ . \end{aligned} \tag{7.12}$$

The junction is a sequence of tokens, interpreted as a unit, providing a way to coordinate two O-groups of some level of complexity. To simplify things, junctions are represented in the dictionary as single tokens. In fact, the lexical analyser would combine tokens when needed. These special tokens are marked with a tag denoting their status.[17]

$$\text{OJunct}(u,v,c) \quad \equiv \quad u = v \wedge \mathsf{OJunct} \in u \wedge \mathsf{OAgr}(c) \in u \ . \tag{7.13}$$

Fundamental stative groups

The fundamental stative group, or O-group, is one of the following.

$$\begin{aligned} \text{O-group}(u,v,p) \quad \equiv \quad & \text{o-det-group}(u,v,p) \vee \\ & \vee \text{i-group}(u,v,p) \vee \\ & \vee \text{o-pron-group}(u,v,p) \vee \\ & \vee \text{o-emp-det-group}(u,v,p) \ . \end{aligned} \tag{7.14}$$

The o-det groups previously described share the same structure, as one can easily see by inspection of the corresponding tables. The common

structure is captured by the following predicate.

$$\begin{aligned} \text{o-det-group}(u,v,p) \quad \equiv \quad & \exists_{[u,v]} y,z,w,t.\exists c.p = \text{-o} \circ c \wedge \\ & \wedge \text{Det}(u,y,c) \wedge \text{Adj}^{\oslash}_{*}(\overrightarrow{y},\overleftarrow{z},c) \wedge \\ & \wedge \lambda\text{-group}(z,w) \wedge \mathsf{Rep}(\text{-o}) \in \overrightarrow{w} \wedge \\ & \wedge \text{Agr}(\overrightarrow{\overrightarrow{w}},t,c) \wedge \text{Adj}^{\ominus}_{*}(\overrightarrow{t},v,c) \ . \end{aligned} \tag{7.15}$$

The determiner is recognised by checking to which class it belongs (section 5.8).

$$\begin{aligned} \text{Det}(u,v,c) \quad \equiv \quad & \exists a,d,l.\text{DetInv}(u,v,l) \vee \text{DetInvCorr}(u,v,a,l) \vee \\ & \vee \text{APron}(u,v,c,l) \vee \text{DetVar}(u,v,c,a,l) \vee \\ & \vee \text{ANum}(u,v,l,d) \ . \end{aligned} \tag{7.16}$$

Furthermore, there can be zero, one, or more adj-groups. An adj-group can be either Slash or Minus, as already explained in section 6.2.6. Adj-groups may be present in any number or they may be absent. This form of occurrence is captured via the following predicates. They unroll the recursive definition with respect to the proximity to the "focus" of the O-group.

$$\begin{aligned} \text{Adj}^{\oslash}_{*}(u,v,c) \quad \equiv \quad & (\exists_{[u,v]} w.\text{Adj-group}(w,v,\oslash,c) \wedge \text{Adj}^{\oslash}_{*}(u,\overleftarrow{w},c)) \vee \\ & \vee u \rightarrowtail v \ . \end{aligned} \tag{7.17}$$

$$\begin{aligned} \text{Adj}^{\ominus}_{*}(u,v,c) \quad \equiv \quad & (\exists_{[u,v]} w.\text{Adj-group}(u,w,\ominus,c) \wedge \text{Adj}^{\oslash}_{*}(\overrightarrow{w},v,c)) \vee \\ & \vee u \rightarrowtail v \ . \end{aligned} \tag{7.18}$$

Agreement checks number (plural vs. singular) and case (accusative vs. nominative).

$$\begin{aligned} \text{Agr}(u,v,c) \quad \equiv \quad & (\mathsf{Rep}(\text{-j}) \in u \wedge \mathsf{Rep}(\text{-n}) \in v \wedge \\ & \quad \wedge \overrightarrow{u} = v \wedge c = \text{-jn}) \vee \\ & \vee (\mathsf{Rep}(\text{-j}) \in u \wedge u = v \wedge c = \text{-j}) \vee \\ & \vee (\mathsf{Rep}(\text{-n}) \in u \wedge u = v \wedge c = \text{-n}) \vee \\ & \vee (u \rightarrowtail v \wedge c = \emptyset) \ . \end{aligned} \tag{7.19}$$

There are various types of determiners that concur to form stative groups (see Table 6.6 and the following). It should be noted that some determiners are members of the A-group, so they are defined in their appropriate section.

$$\begin{aligned} \text{DetInv}(u,v,l) \quad \equiv \quad & (\mathsf{Rep}(l) \in u \wedge \mathsf{Invar} \in u \wedge u = v) \vee \\ & \vee (u \rightarrowtail v \wedge l = \square) \ . \end{aligned} \tag{7.20}$$

$$\text{DetInvCorr}(u,v,a,l) \equiv \mathsf{Rep}(l) \in u \wedge \mathsf{AdjCorr} \in u \wedge \wedge \mathsf{Rep}(a) \in v \wedge \mathsf{InvAdjCorr} \in v \wedge \wedge \overrightarrow{u} = v \ . \quad (7.21)$$

$$\text{DetVar}(u,v,c,a,l) \equiv \mathsf{Rep}(l) \in u \wedge \mathsf{AdjCorr} \in u \wedge \wedge \mathsf{Rep}(a) \in \overrightarrow{u} \wedge \mathsf{VarAdjCorr} \in \overrightarrow{u} \wedge \wedge \text{Agr}(\overrightarrow{\overrightarrow{u}}, v, c) \ . \quad (7.22)$$

Apart from det-groups, an O-group may be an infinitive. In this case, it is called i-group (Table 6.12).

$$\text{i-group}(u,v,p) \equiv \mathsf{Rep}(\text{-i}) \in v \wedge \lambda\text{-group}(u, \overleftarrow{v}) \wedge p = \text{-i} \ . \quad (7.23)$$

Moreover, an O-group may be a pronoun. As said, these O-groups have a similar structure – but still different – from det-groups.

$$\begin{aligned}\text{o-pron-group}(u,v,p) \equiv\ & \exists n.\mathsf{Pron} \in u \wedge p = n \wedge \\ & \wedge ((u = v \wedge n = \emptyset) \vee \\ & \quad \vee (\exists l,d.\text{ANum}(\overrightarrow{u}, v, l, d) \wedge n = \emptyset) \vee \\ & \quad \vee (\mathsf{Rep}(\text{-n}) \in \overrightarrow{u} \wedge \overrightarrow{u} = v \wedge n = \text{-n}) \vee \\ & \quad \vee (\exists l,d.\mathsf{Rep}(\text{-n}) \in \overrightarrow{u} \wedge n = \text{-n} \wedge \\ & \qquad \wedge \text{ANum}(\overrightarrow{\overrightarrow{u}}, v, l, d))) \ .\end{aligned} \quad (7.24)$$

Finally, an O-group may be a trace group, properly called an empty determiner group, or o-emp-det-group for short.

$$\begin{aligned}\text{o-emp-det-group}(u,v,p) \equiv\ & \exists_{[u,v]} y,z,w.\exists c,a,l,l'.p = c \wedge \\ & \wedge \text{EmpDet}(u,y,c) \wedge \text{APron}(\overrightarrow{y}, z, c, l) \wedge \\ & \wedge \text{Adj}^{\emptyset}_{*}(\overrightarrow{z}, \overleftarrow{w}, c) \wedge \text{ANum}(w, v, l', a) \ .\end{aligned} \quad (7.25)$$

An empty determiner is either a DetInv (Table 6.14) or a DetVar (Table 6.14).

$$\text{EmpDet}(u,v,c) \equiv \exists a,l.\text{DetInv}(u,v,l) \vee \text{DetVar}(u,v,c,a,l) \ . \quad (7.26)$$

An o-group has a case. The following predicate matches exactly nominative groups.

$$\begin{aligned}\text{nom-stative-group}(x,y) \equiv\ & \text{stative-group}(x,y,\emptyset) \vee \\ & \vee \text{stative-group}(x,y,\text{-j}) \ .\end{aligned} \quad (7.27)$$

In contrast, the following predicate matches exactly accusative groups.

$$\text{acc-stative-group}(x,y) \equiv \exists a.\text{stative-group}(x,y,a \circ \text{-n}) \ . \quad (7.28)$$

What has been called a prepositional clause is nothing else than a stative group with a preposition which qualifies its role in the phrase, as defined by the following predicate which extracts the preposition in the p variable.

$$\begin{aligned}\text{prep-nom-stative-group}(x,y,p) \equiv\ & \exists a. \\ & \text{stative-group}(x,y,p \lozenge a) \wedge \\ & \wedge (a = \epsilon \vee a = \text{-j}) \ .\end{aligned} \quad (7.29)$$

$$\begin{aligned}\text{prep-acc-stative-group}(x,y,p) \equiv\ & \exists a. \\ & \text{stative-group}(x,y,p \lozenge a \circ \text{-n}) \ .\end{aligned} \quad (7.30)$$

7.4.4 A-predicates

A-predicates are deputed to recognise adjunctives, i.e., the modifiers of O-groups. The prominent members of this family are the adjectives. But, as already explained, other linguistic elements may be regarded as A-modifiers (section 6.2.4).

As for the case of O-groups, this group is characterised by the A-group predicate, which defines its fundamental structure. But, A-groups may be composed, so the need for a pair of predicates that take care to model the composition of A-groups.

$$\begin{aligned}\text{Adj-group}(u,v,s,c) \equiv\ & \text{Adj1-group}(u,v,s,c) \vee \\ & \vee (\exists_{(u,v)} x.\exists c.\mathsf{APreJunct} \in u \wedge \mathsf{Rep}(c) \in u \wedge \\ & \quad \wedge \text{Adj1-group}(\overrightarrow{u}, \overleftarrow{x}, s, c) \wedge \\ & \quad \wedge \mathsf{APostJunct}(c) \in x \wedge \\ & \quad \wedge \text{Adj1-group}(\overrightarrow{x}, v, s, c)) \ .\end{aligned} \quad (7.31)$$

$$\begin{aligned}\text{Adj1-group}(u,v,s,c) \equiv\ & \text{A-group}(u,v,s,c) \vee \\ & \vee (\exists_{(u,v)} x.\text{Adj1-group}(u, \overleftarrow{x}, s, c) \wedge \\ & \quad \wedge \mathsf{AJunct} \in x \wedge \\ & \quad \wedge \text{Adj1-group}(\overrightarrow{x}, v, s, c)) \ .\end{aligned} \quad (7.32)$$

An A-group comes in two forms: either it is an adjective, or it may be a phrase introducted by an appropriate junction. Furthermore, a preposi-

tional clause modifying a stative group *is* an adjunctive.

$$\text{A-group}(u,v,s,c) \equiv \text{A1-group}(u,v,c) \vee \text{A2-group}(u,v,c) \vee (\text{7.33})$$
$$\vee (c = \emptyset \wedge \text{AJoined}(u,v)) \vee$$
$$\vee (c = \emptyset \wedge \text{circ-group}(u,v)) \ .$$

The two adjective forms of the A-group has been analysed above. They are coded by the following predicates.

$$\text{A1-group}(u,v,c) \equiv \exists_{[u,v]} x,y,w.\exists l,l',d.\text{EAdj-group}(w,v,c) \wedge (\text{7.34})$$
$$\wedge\ \text{e-group}^*(\overrightarrow{y},\overleftarrow{w}) \wedge \text{ANum}(x,y,l,d) \wedge$$
$$\wedge\ \text{APron}(u,\overleftarrow{x},c,l') \ .$$

$$\text{A2-group}(u,v,c) \equiv \exists_{[u,v]} x,y,w.\exists l,l',d.\text{EAdj-group}(w,v,c) \wedge (\text{7.35})$$
$$\wedge\ \text{e-group}^*(\overrightarrow{y},\overleftarrow{w}) \wedge \text{APron}(x,y,c,l) \wedge$$
$$\wedge\ \text{ANum}(u,\overleftarrow{x},l',d) \ .$$

Last, but very important, an Adj-group may be a phrase preceded by a junctive of the right class.

$$\text{AJoined}(u,v) \equiv \mathsf{AHJunct} \in u \wedge \text{phrase-group}(\overrightarrow{u},v) \ . \quad (\text{7.36})$$

Both the A1-groups and the A2-groups have a common kernel, identified by the EAdj-group, i.e., elementary adjectival groups (e-adj-groups), already explained through Table 6.16.

$$\text{EAdj-group}(u,v,c) \equiv \exists_{u,v} w.\mathsf{Rep}(\text{-a}) \in w \wedge \quad (\text{7.37})$$
$$\wedge\ \lambda\text{-group}(u,\overleftarrow{w}) \wedge \text{Agr}(\overrightarrow{w},v,c) \ .$$

The EAdj-groups may be possibly modified by one or more e-groups (see section 6.2.2).

$$\text{e-group}^*(u,v) \equiv (\exists_{[u,v]} w.\text{e-group}(w,v) \wedge \text{e-group}^*(u,\overleftarrow{w})) \vee (\text{7.38})$$
$$\vee\ u \rightarrowtail v \ .$$

The APron is a pronoun which agrees with the O-group the whole Adj-group modifies (Table 6.15).

$$\text{APron}(u,v,c,l) \equiv \mathsf{Rep}(l) \in u \wedge \mathsf{Pron} \in u \wedge \quad (\text{7.39})$$
$$\wedge\ \mathsf{Rep}(\text{-a}) \in \overrightarrow{u} \wedge \text{Agr}(\overrightarrow{\overrightarrow{u}},v,c) \ .$$

The ANum is a particular numerical predicate explained in section 6.2.5. As explained before, numbers and numerical expression form an intricate

subject (section 5.8.3). There is no reason to really code these forms into the grammar, since they affect only the way a sequence of token is interpreted as a unit. Specifically, a numerical object is nothing else than a single token recognised during the lexical analysis phase. Hence, I assume to have a clever dictionary.[18] The dictionary calculates whether a sequence of tokens can be thought as a numerical expression, so I need to cope only with the presence of determiners that may affect the way the adtree is built.

$$\text{ANum}(u,v,l,d) \equiv (\exists c.\mathsf{NumDet} \in u \wedge \mathsf{NumAtom}(c) \in u \wedge \quad (\text{7.40})$$
$$\wedge\ \mathsf{Num} \in v \wedge \mathsf{NumAtom}(c) \in v \wedge$$
$$\wedge\ \mathsf{Rep}(d) \in u \wedge \mathsf{Rep}(l) \in v \wedge \overrightarrow{u} = v) \vee$$
$$\vee\ (\mathsf{Num} \in u \wedge u = v \wedge \mathsf{Rep}(l) \in u \wedge d = \emptyset) \ .$$

7.4.5 E-predicates

E-groups denote verbal modifiers. A slight complication is that, sometimes, E-groups are required not to admit the accusative form. For this reason, also e-groups have been defined (section 6.2.2).

Since E-groups may be composed, the definition of the E-group predicate is suspiciosly similar to the ones of O-groups and A-groups. In fact, an E-group may be either simple or the composition of a pair of E-groups, properly joined. A simple E-group is either an E-S-group or a circumstantial group.

$$\text{E-group}(u,v) \equiv \text{E1-group}(u,v) \vee \quad (\text{7.41})$$
$$\vee\ (\exists_{(u,v)} x.\exists c.\mathsf{EPreJunct} \in u \wedge \mathsf{Rep}(c) \in u \wedge$$
$$\wedge\ \text{E1-group}(\overrightarrow{u},\overleftarrow{x}) \wedge$$
$$\wedge\ \mathsf{EPostJunct}(c) \in x \wedge$$
$$\wedge\ \text{E1-group}(\overrightarrow{x},v)) \ .$$

$$\text{E1-group}(u,v) \equiv \text{E-S-group}(u,v) \vee \text{circ-group}(u,v) \vee \quad (\text{7.42})$$
$$\vee\ (\exists_{(u,v)} x.\text{E1-group}(u,\overleftarrow{x}) \wedge$$
$$\wedge\ \mathsf{EJunct} \in x \wedge$$
$$\wedge\ \text{E1-group}(\overrightarrow{x},v)) \ .$$

An E-S-group, where {S} stands for 'simple', is composed by an optional preposition, followed by a λ-subgroup, and concluded by *-e* or by

-en. Also, it may be a prepositional clause.

$$\begin{aligned}\text{E-S-group}(u,v) \;\equiv\; & \text{e-S-group}(u,v) \vee \\ & \vee \text{circ-group}(u,v) \vee \\ & \vee ((\mathsf{Rep}\text{(-e)} \in \overleftarrow{v} \wedge \lambda\text{-group}(u, \overleftarrow{\overleftarrow{v}}) \wedge \\ & \quad \wedge \mathsf{Rep}\text{(-n)} \in v) \;.\end{aligned} \tag{7.43}$$

A circ-group models prepositional clauses modfiying I-groups.

$$\begin{aligned}\text{circ-group}(u,v) \;\equiv\; & \exists p.\text{pre-nom-stative-group}(u,v,p) \vee \\ & \vee \text{pre-acc-stative-group}(u,v,p) \;.\end{aligned} \tag{7.44}$$

An e-group is similar to an E-group except that it does not admit the signature of the accusative case *-n*. Its recognising predicates are defined accordingly.

$$\begin{aligned}\text{e-group}(u,v,s,c) \;\equiv\; & \text{e-S-group}(u,v) \vee \text{circ-group}(u,v) \vee \\ & \vee (\exists_{(u,v)} x.\text{e-group}(u, \overleftarrow{x}) \wedge \\ & \quad \wedge \mathsf{EJunct} \in x \wedge \\ & \quad \wedge \text{e-group}(\overrightarrow{x}, v)) \;.\end{aligned} \tag{7.45}$$

$$\begin{aligned}\text{e-S-group}(u,v) \;\equiv\; & (\mathsf{Rep}\text{(-e)} \in v \wedge \lambda\text{-group}(u, \overleftarrow{v})) \vee \\ & \vee (\mathsf{EDetEnv} \in u \wedge \mathsf{Rep}\text{(-e)} \in v \wedge \\ & \quad \wedge \lambda\text{-group}(\overrightarrow{u}, \overleftarrow{v})) \vee \\ & \vee \text{EJoined}(u,v) \;.\end{aligned} \tag{7.46}$$

Last, but very important, an E-group or an e-group may be a phrase preceded by a junctive of the right class.

$$\text{EJoined}(u,v,c) \;\equiv\; \mathsf{EHJunct} \in u \wedge \text{phrase-group}(\overrightarrow{u}, v) \;. \tag{7.47}$$

Let me comment briefly this linguistic discover, that was driven rightly by the formal model: there are phrases preceded by a junctive that are 'declassed' to clauses joined circumstantially or adjectivally to their 'real' phrase. This discover was what lead me to re-define the levels of analysis concerning NLs.[19]

7.4.6 The phrase predicates

A phrase is composed by a verb with the structure induced by its valence as well as its modifiers.

$$\begin{aligned}\text{phrase-group}(u,v) \;\equiv\; & \text{Verb}_0(u,v) \vee \text{Verb}_1(u,v) \vee \\ & \vee \text{Verb}_2(u,v) \vee \text{Verb}_3(u,v) \;.\end{aligned} \tag{7.48}$$

The structure of a zerovalent verb is a I-group possibly surrounded externally by E-groups.[20]

$$\begin{aligned}\text{Verb}_0(u,v) \;\equiv\; & (\exists_{[u,v]} x,y.\exists c,d. \\ & \text{E}^*\text{-group}(u, \overleftarrow{x}) \wedge \\ & \wedge \text{I-group}(x,y,0,d,c) \wedge \\ & \wedge \text{E}^*\text{-group}(\overrightarrow{y}, v)) \;.\end{aligned} \tag{7.49}$$

The structure of a monovalent verb is a subject and a I-group possibly surrounded by E-groups. The following predicate is incomplete for conciseness and clarity: the omitted part is simply obtained by moving the subject (nom-stative-group) after the I-group. The omitted part takes care of the previously explained mirroring phenomenon.

$$\begin{aligned}\text{Verb}_1(u,v) \;\equiv\; & (\exists_{[u,v]} x,y,s_1,s_2.\exists c,d. \\ & \text{E}^*\text{-group}(u, \overleftarrow{s_1}) \wedge \\ & \wedge \text{nom-stative-group}(s_1,s_2) \wedge \\ & \wedge \text{E}^*\text{-group}(\overrightarrow{s_1}, \overleftarrow{x}) \wedge \\ & \wedge \text{I-group}(x,y,1,d,c) \wedge \\ & \wedge \text{E}^*\text{-group}(\overrightarrow{y}, v)) \vee \ldots \;.\end{aligned} \tag{7.50}$$

The structure of a bivalent verb is a subject, a I-group and an object, possibly surrounded by E-groups. Again, mirroring is hidden.

$$\begin{aligned}\text{Verb}_2(u,v) \;\equiv\; & (\exists_{[u,v]} x,y,s_1,s_2,o_1,o_2.\exists c,d. \\ & \text{E}^*\text{-group}(u, \overleftarrow{s_1}) \wedge \\ & \wedge \text{nom-stative-group}(s_1,s_2) \wedge \\ & \wedge \text{E}^*\text{-group}(\overrightarrow{s_1}, \overleftarrow{x}) \wedge \\ & \wedge \text{I-group}(x,y,2,d,c) \wedge \\ & \wedge \text{E}^*\text{-group}(\overrightarrow{y}, \overleftarrow{o_1}) \wedge \\ & \wedge \text{acc-stative-group}(o_1,o_2) \wedge \\ & \wedge \text{E}^*\text{-group}(\overrightarrow{o_2}, v)) \vee \ldots \;.\end{aligned} \tag{7.51}$$

The structure of a trivalent verb is: a subject, a I-group, an object, and a

dative, possibly surrounded by E-groups.

$$\begin{aligned}\text{Verb}_3(u,v) \equiv\ & (\exists_{[u,v]} x,y,s_1,s_2,o_1,o_2,d_1,d_2.\exists c,d,e. \\ & \text{E}^*\text{-group}(u,\overleftarrow{s_1}) \wedge \\ & \wedge\ \text{nom-stative-group}(s_1,s_2) \wedge \\ & \wedge\ \text{E}^*\text{-group}(\overrightarrow{s_1},\overleftarrow{x}) \wedge \\ & \wedge\ \text{I-group}(x,y,3,d,c) \wedge \\ & \wedge\ \text{E}^*\text{-group}(\overrightarrow{y},\overleftarrow{o_1}) \wedge \\ & \wedge\ \text{acc-stative-group}(o_1,o_2) \wedge \\ & \wedge\ \text{E}^*\text{-group}(\overrightarrow{o_2},\overleftarrow{d_1}) \wedge \\ & \wedge\ \text{prep-nom-stative-group}(d_1,d_2,\text{al} \Diamond e) \wedge \\ & \wedge\ \text{E}^*\text{-group}(\overrightarrow{d_2},v)) \vee \ldots \ .\end{aligned} \tag{7.52}$$

Every E*-group can be an empty sequence of E-groups, each one modifying the verb.

$$\begin{aligned}\text{E}^*\text{-group}(u,v) \equiv\ & (\exists_{[u,v]} w.\text{E-group}(w,v) \wedge \text{E}^*\text{-group}(u,\overleftarrow{w})) \vee \\ & \vee\ u \rightarrowtail v \ .\end{aligned} \tag{7.53}$$

7.4.7 The sentence predicates

A sentence is composed by a single phrase or by two subsentences joined together (see Table 5.17).

$$\begin{aligned}\text{sentence-group}(u,v) \equiv\ & \text{phrase-group}(u,v) \vee \\ & \vee\ (\exists_{(u,v)} x.\text{sentence-group}(u,\overleftarrow{x}) \wedge \\ & \wedge\ \textsf{SJunct} \in x \wedge \\ & \wedge\ \text{sentence-group}(\overrightarrow{x},v)) \ .\end{aligned} \tag{7.54}$$

7.4.8 The text predicates

The whole text is a sequence of sentences, each one joined to the following by means of a delimiter (e.g., questions are delimited by question marks, see for instance Figure 5.13).

$$\begin{aligned}\text{text-group}(u,v) \equiv\ & \text{sentence-group}(u,v) \vee \\ & \vee\ (\exists_{(u,v)} x.\text{sentence-group}(u,\overleftarrow{x}) \wedge \\ & \wedge\ \textsf{SentenceDelimiter} \in x \wedge \\ & \wedge\ \text{text-group}(\overrightarrow{x},v)) \ .\end{aligned} \tag{7.55}$$

The parsing machine recognises a text also if its last sentence is not delimited. Note that the empty text is a valid text.

$$\begin{aligned}\text{init-group}(u,v) \equiv\ & u \rightarrowtail v \vee \text{text-group}(u,v) \vee \\ & \vee\ (\text{text-group}(u,\overleftarrow{v}) \wedge \textsf{SentenceDelimiter} \in v) \ .\end{aligned} \tag{7.56}$$

Let me conclude this section with an important remark. The predicate 56 could be named appropriately Esperanto-group, as *all* the grammar of this QNL is coded through exactly 56 formulae, technically speaking. This means that predicate 56 is satisfiable if and only if its models are true, i.e., they are series of tokens (texts) written in standard Esperanto. In other words, predicate 56 checks if an input text is an Esperanto text or not.

7.5 Parsing rules

The predicates introduced so far can be used to validate a text. In fact, asking whether a given token list satisfies the init-group predicate, is equivalent to ask whether the token list represents a correct text written in Esperanto.

Nevertheless, the previous predicates do not represent the text in a useful format. They do not help the understanding, nor they facilitate the translation of the text into another language, nor any other transformation a user may want to apply on the text itself.

In order to have a suitable representation a parsing mechanism is needed. This would transform a text – represented as a list of tokens – into an adtree. Then, the adtree may be used to analyse, to transform or to translate the text. As explained in the previous chapters, the structure of the text is made evident in the tree, and so structural transformations are easily described as acting on the appropriate adtrees.

This section shows the formal parsing rules that construct adtrees from texts. These rules have the format previously described and they are intended to operate in the parsing environment introduced at the beginning of this chapter. The rules are divided according to their function, roughly following the guidelines of the previous chapters and the taxonomy of predicates made in the previous section.[21]

7.5.1 Structural rules

Structural rules are intended to aid the derivation process but they do not parse the current text.

The Struct_1 rule destroys empty ranges. It should be noted that λ-ranges should not be canceled when empty, since empty λ-ranges have still

a meaning even if they contain no token.

$$\dfrac{\begin{array}{l}\triangleright\, w(x,y)\\ \triangleright\, T[a]\\ \triangleright\, x \rightarrowtail y \wedge w \neq \lambda\text{-range}\end{array}}{\begin{array}{l}\triangleright\, \{-w(x,y)\}\\ \triangleright\, T[a]\end{array}}\ \text{Struct}_1 \tag{7.57}$$

The Struct_2 rule merges adiacent ranges of the same kind.[22]

$$\dfrac{\begin{array}{l}\triangleright\, w(x,y) \wedge w(\overrightarrow{y},z)\\ \triangleright\, T[a]\\ \triangleright\, \top\end{array}}{\begin{array}{l}\triangleright\, \{+w(x,z), -w(x,y), -w(\overrightarrow{y},z)\}\\ \triangleright\, T[a]\end{array}}\ \text{Struct}_2 \tag{7.58}$$

The Struct_3 rules destroy ranges with an invalid bound.

$$\dfrac{\begin{array}{l}\triangleright\, w(\emptyset,x)\\ \triangleright\, T[a]\\ \triangleright\, w \neq \lambda\text{-range}\end{array}}{\begin{array}{l}\triangleright\, \{-w(\emptyset,x)\}\\ \triangleright\, T[a]\end{array}}\ \text{Struct}_3 \tag{7.59}$$

$$\dfrac{\begin{array}{l}\triangleright\, w(x,\emptyset)\\ \triangleright\, T[a]\\ \triangleright\, w \neq \lambda\text{-range}\end{array}}{\begin{array}{l}\triangleright\, \{-w(x,\emptyset)\}\\ \triangleright\, T[a]\end{array}}\ \text{Struct}_3 \tag{7.60}$$

7.5.2 The parsing of λ-subgroups

The whole group is parsed by the following rules. The first one parses the atom and the specifier when present, the second one parses the atom when the specifier is absent and the third one parses the empty group (see section

6.1).

$$\dfrac{\begin{array}{lll}\triangleright\, \lambda\text{-range}(u,v) & \triangleright\, \{\lambda\text{-atom-range}(\overrightarrow{w},v)\} & \triangleright\, \{\lambda\text{-spec-range}(u,w)\}\\ \triangleright\, \emptyset & \triangleright\, \emptyset & \triangleright\, \emptyset\\ & \vdots\,\pi_{\text{atom}} & \vdots\,\pi_{\text{spec}}\\ \triangleright\, w \in \lambda\text{-range}(u,v) \wedge & \triangleright\, \emptyset & \triangleright\, \emptyset\\ \quad \wedge\, \lambda\text{-atom-group}(\overrightarrow{w},v,n,a) & \triangleright\, B & \triangleright\, A_{[\mathsf{Rep}(l),\mathsf{Sign}(\varnothing),\mathsf{Dir}(\rightarrow)]}\end{array}}{\begin{array}{l}\triangleright\, \{-\lambda\text{-range}(u,v)\}\\ \triangleright\, \overset{\varnothing}{\underset{l}{\rightarrow}}\,(A,B)\end{array}}\ \lambda_1 \tag{7.61}$$

$$\dfrac{\begin{array}{l}\triangleright\, \lambda\text{-range}(u,v)\\ \triangleright\, \emptyset\\ \triangleright\, \lambda\text{-atom-group}(u,v,n,a)\end{array}}{\begin{array}{l}\triangleright\, \{+\lambda\text{-atom-range}(u,v), -\lambda\text{-range}(u,v)\}\\ \triangleright\, \emptyset\end{array}}\ \lambda_2 \tag{7.62}$$

The empty range is easily parsed by the following rule. It is noteworthy that the structural rules have been conceived not to cancel this range.

$$\dfrac{\begin{array}{l}\triangleright\, \lambda\text{-range}(u,v)\\ \triangleright\, \emptyset\\ \triangleright\, u \rightarrowtail v\end{array}}{\begin{array}{l}\triangleright\, \{-\lambda\text{-range}(u,v)\}\\ \triangleright\, \square\end{array}}\ \lambda_3 \tag{7.63}$$

The specifier is parsed by constructing an adtree whose node above the root is the fundamental adposition. The following pair of rules do the job: the former treats the case when the fundamental adposition is absent, while the latter deals with the case in which the adposition is present.

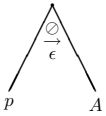
$$\dfrac{\begin{array}{ll}\triangleright\, \{\lambda\text{-prefix-range}(u,\overleftarrow{v})\} & \\ \triangleright\, k & \\ \triangleright\, \lambda\text{-spec-range}(u,v) & \\ \triangleright\, \emptyset & \vdots\,\pi_{\text{prefix}}\\ \triangleright\, \lambda\text{-spec-group}(u,v,\emptyset,a) \wedge & \triangleright\, \emptyset\\ \quad \wedge\, \mathsf{Rep}(k) \in v & \triangleright\, P\end{array}}{\begin{array}{l}\triangleright\, \{-\lambda\text{-spec-range}(u,v)\}\\ \triangleright\, P_{[\mathsf{Rep}(\epsilon),\mathsf{Sign}(\varnothing),\mathsf{Dir}(\rightarrow)]}\end{array}}\ \lambda\text{-spec}_1 \tag{7.64}$$

$$\dfrac{\begin{array}{ll}\triangleright\ \lambda\text{-spec-range}(u,v) & \triangleright\ \{\lambda\text{-prefix-range}(u,\overleftarrow{v})\}\\ \triangleright\ \emptyset & \triangleright\ k\\ \triangleright\ \lambda\text{-spec-group}(u,v,l,a)\ \wedge & \vdots\pi_{\text{prefix}}\\ \quad\wedge\ l\neq\emptyset\wedge\mathsf{Rep}(k)\in\overleftarrow{v} & \triangleright\ \emptyset\\ & \triangleright\ P\end{array}}{\begin{array}{l}\triangleright\ \{-\lambda\text{-spec-range}(u,v)\}\\ \triangleright\ P_{[\mathsf{Rep}(l),\mathsf{Sign}(\oslash),\mathsf{Dir}(\rightarrow)]}\end{array}}\ \lambda\text{-spec}_2 \tag{7.65}$$

The following rule parses the prefix. The derivation is concluded when the range becomes empty and, thus, it gets discarded by a structural rule.

$$\dfrac{\begin{array}{l}\triangleright\ \lambda\text{-prefix-range}(u,v)\\ \triangleright\ A\\ \triangleright\ \mathsf{Rep}(p)\in v\end{array}}{\triangleright\ \{+\lambda\text{-prefix-range}(u,\overleftarrow{v}),-\lambda\text{-prefix-range}(u,v)\}}\ \lambda\text{-prefix} \tag{7.66}$$

$\triangleright$



The atom is parsed by combining its lexeme with the suffixes. There are no special suffixes with respect to the adtree construction, since they have already been checked by a previous rule.

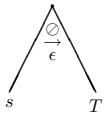
$$\dfrac{\begin{array}{l}\triangleright\ \lambda\text{-atom-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \mathsf{Rep}(l)\in u\end{array}}{\begin{array}{l}\triangleright\ \{+\lambda\text{-suffix-range}(\overrightarrow{u},v),-\lambda\text{-atom-range}(u,v)\}\\ \triangleright\ l\end{array}}\ \lambda\text{-atom} \tag{7.67}$$

The suffixes are recursively parsed from left to right by a sequence of applications of the following rule. When the range becomes empty, a struc-

tural rule takes care of concluding the derivation (section 6.3).

$$\dfrac{\begin{array}{l}\triangleright\ \lambda\text{-suffix-range}(u,v)\\ \triangleright\ T\\ \triangleright\ \mathsf{Rep}(s)\in u\end{array}}{\triangleright\ \{+\lambda\text{-suffix-range}(\overrightarrow{u},v),-\lambda\text{-suffix-range}(u,v)\}}\ \lambda\text{-suffix} \tag{7.68}$$

$\triangleright$



7.5.3 The I-group

The I-group describes a verb. Thus, this set of rules parse an I-range, corresponding to a sequence of tokens satisfying the I-group predicate.

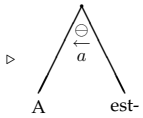
The general I-range gets parsed by the following rule. The extraction of the verbal final is left to the rules that build the adtree for a phrase.

$$\dfrac{\begin{array}{l}\triangleright\ \text{I-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \lambda\text{-group}(u,\overleftarrow{v})\end{array}}{\begin{array}{l}\triangleright\ \{+\lambda\text{-range}(u,\overleftarrow{v}),-\text{I-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{I}_1 \tag{7.69}$$

As said before, the verb *esti* is an exception (section 5.5.5), so the following rules apply when the it is used in conjunction with an adjective, a circumstantial or a nominal stative, respectively.

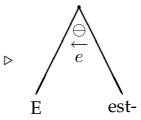
$$\dfrac{\begin{array}{ll}\triangleright\ \text{I-range}(u,v) & \triangleright\ \{\text{Adj-range}(\overrightarrow{\overrightarrow{u}},v)\}\\ \triangleright\ \emptyset & \triangleright\ \emptyset\\ \triangleright\ \text{Adj-group}(\overrightarrow{\overrightarrow{u}},v,\ominus,c) & \vdots\pi_{\text{Adj}}\\ & \triangleright\ \emptyset\\ & \triangleright\ A_{[\mathsf{Rep}(a),\mathsf{Sign}(\ominus),\mathsf{Dir}(\leftarrow)]}\end{array}}{\triangleright\ \{-\text{I-range}(u,v)\}}\ \text{I}_2 \tag{7.70}$$

$\triangleright$



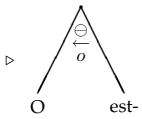
(7.71)

$$\dfrac{\begin{array}{ll}\triangleright\ \text{I-range}(u,v) & \vdots\pi_{\text{e}}\\ \triangleright\ \emptyset & \triangleright\ \{\text{E-range}(\overrightarrow{u},v)\}\\ \triangleright\ \text{e-group}(\overrightarrow{u},v) & \triangleright\ \emptyset\\ & \triangleright\ E_{[\text{Rep}(e),\text{Sign}(\ominus),\text{Dir}(\leftarrow)]}\end{array}}{\triangleright\ \{-\text{I-range}(u,v)\}}\ \text{I}_3$$



(7.72)

$$\dfrac{\begin{array}{ll}\triangleright\ \text{I-range}(u,v) & \vdots\pi_{\text{stative}}\\ \triangleright\ \emptyset & \triangleright\ \{\text{stative-range}(\overrightarrow{u},v)\}\\ \triangleright\ \text{nom-stative-group}(\overrightarrow{u},v) & \triangleright\ \emptyset\\ & \triangleright\ O_{[\text{Rep}(o),\text{Sign}(\ominus),\text{Dir}(\leftarrow)]}\end{array}}{\triangleright\ \{-\text{I-range}(u,v)\}}\ \text{I}_4$$



7.5.4 The O-group

The stative group contains the rules deputed to parse stative lexemes and their modifiers. They can be applied when a stative-range is present in the current state.

The whole stative range is parsed by the following rules, strictly following the definition of the stative-group predicate, i.e., *all* presented in section 6.2.3 and its subsection is represented here.

(7.73)

$$\dfrac{\begin{array}{l}\triangleright\ \text{stative-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{stative1-group}(u,v,p)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{stative1-range}(u,v),-\text{stative-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{stative}_1$$

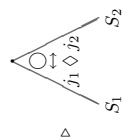
The second rule is quite complex. The generated adtree has s as sign and d as direction. The first case copes with examples like

- (I'm going to the sea) both with the dogs and by car.

The sign and the direction of the hook are the same as seen for circumstantials, while its representation is the composition of the prepositions. [23]

(7.74)

$$\dfrac{\begin{array}{ll}\triangleright\ \text{stative-range}(u,v) & \triangleright\ \{\text{stative1-range}(\overrightarrow{u},\overleftarrow{x})\}\\ \triangleright\ \emptyset & \triangleright\ \emptyset\\ \triangleright\ x\in\text{stative-range}(u,v)\wedge & \vdots\pi_{\text{stative}}\\ \quad\wedge\ \text{OPreJunct}\in u\wedge & \triangleright\ \{\text{stative1-range}(\overrightarrow{u},\overleftarrow{x})\}\\ \quad\wedge\ \text{Rep}(j_1)\in u\wedge & \triangleright\ \emptyset\\ \quad\wedge\ \text{OPostJunct}(j_1)\in x\wedge & \vdots\pi_{\text{stative}}\\ \quad\wedge\ \text{Rep}(j_2)\in x\wedge & \triangleright\ \emptyset\\ \quad\wedge\ \text{Sign}(s)\in x\wedge & \triangleright\ S_1{}_{[\text{Rep}(p_1\Diamond c_1),\text{Sign}(s_1),\text{Dir}(d_1)]}\\ \quad\wedge\ \text{Dir}(d)\in x & \triangleright\ S_2{}_{[\text{Rep}(p_2\Diamond c_2),\text{Sign}(s_2),\text{Dir}(d_2)]}\end{array}}{\begin{array}{l}\triangleright\ \{-\text{stative-range}(u,v)\}\\ \qquad\qquad [\text{Rep}(p_1\Diamond p_2\Diamond(c_1\sharp_{j_2}c_2)),\text{Sign}(\oslash),\text{Dir}(\rightarrow)]\end{array}}\ \text{stative}_2$$



When there is no preposition, the whole group may act as a stative, so its hook is changed accordingly.

$$\frac{\begin{array}{l}\triangleright\ \text{stative-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ x \in \text{stative-range}(u,v) \wedge\\ \quad \wedge\ \mathsf{OPreJunct} \in u \wedge\\ \quad \wedge\ \mathsf{Rep}(j_1) \in u \wedge\\ \quad \wedge\ \mathsf{OPostJunct}(j_1) \in x \wedge\\ \quad \wedge\ \mathsf{Rep}(j_2) \in x \wedge\\ \quad \wedge\ \mathsf{Sign}(s) \in x \wedge\\ \quad \wedge\ \mathsf{Dir}(d) \in x\end{array} \quad \begin{array}{l}\triangleright\ \{\text{nom-stative-range}(\overrightarrow{u},\overleftarrow{x})\}\\ \triangleright\ \emptyset\\ \quad\quad \vdots \pi_{\text{stative}}\\ \triangleright\ \emptyset\\ \triangleright\ S_1[\mathsf{Rep}(c_1),\mathsf{Sign}(s_1),\mathsf{Dir}(d_1)]\end{array} \quad \begin{array}{l}\triangleright\ \{\text{nom-stative-range}(\overrightarrow{x},v)\}\\ \triangleright\ \emptyset\\ \quad\quad \vdots \pi_{\text{stative}}\\ \triangleright\ \emptyset\\ \triangleright\ S_2[\mathsf{Rep}(c_2),\mathsf{Sign}(s_2),\mathsf{Dir}(d_2)]\end{array}}{\begin{array}{l}\triangleright\ \{-\text{stative-range}(u,v)\}\\ \triangleright\ \bigl(S_1 \overset{\bigcirc}{\leftrightarrow}_{j_1 \Diamond j_2} S_2\bigr)[\mathsf{Rep}(c_1 \odot_{j_2} c_2),\mathsf{Sign}(\oplus),\mathsf{Dir}(\rightarrow)]\end{array}}\ \text{stative}_3 \tag{7.75}$$

$$\frac{\begin{array}{l}\triangleright\ \text{stative-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ x \in \text{stative-range}(u,v) \wedge\\ \quad \wedge\ \mathsf{OPreJunct} \in u \wedge\\ \quad \wedge\ \mathsf{Rep}(j_1) \in u \wedge\\ \quad \wedge\ \mathsf{OPostJunct}(j_1) \in x \wedge\\ \quad \wedge\ \mathsf{Rep}(j_2) \in x \wedge\\ \quad \wedge\ \mathsf{Sign}(s) \in x \wedge\\ \quad \wedge\ \mathsf{Dir}(d) \in x\end{array} \quad \begin{array}{l}\triangleright\ \{\text{acc-stative-range}(\overrightarrow{u},\overleftarrow{x})\}\\ \triangleright\ \emptyset\\ \quad\quad \vdots \pi_{\text{stative}}\\ \triangleright\ \emptyset\\ \triangleright\ S_1[\mathsf{Rep}(c_1),\mathsf{Sign}(s_1),\mathsf{Dir}(d_1)]\end{array} \quad \begin{array}{l}\triangleright\ \{\text{acc-stative-range}(\overrightarrow{x},v)\}\\ \triangleright\ \emptyset\\ \quad\quad \vdots \pi_{\text{stative}}\\ \triangleright\ \emptyset\\ \triangleright\ S_2[\mathsf{Rep}(c_2),\mathsf{Sign}(s_2),\mathsf{Dir}(d_2)]\end{array}}{\begin{array}{l}\triangleright\ \{-\text{stative-range}(u,v)\}\\ \triangleright\ \bigl(S_1 \overset{\bigcirc}{\leftrightarrow}_{j_1 \Diamond j_2} S_2\bigr)[\mathsf{Rep}(c_1 \odot_{j_2} c_2),\mathsf{Sign}(\ominus),\mathsf{Dir}(\leftarrow)]\end{array}}\ \text{stative}_4 \tag{7.76}$$

The stative1-range is very similar to the stative range: the only difference is that the composing junction lies only in the middle of the group. So,

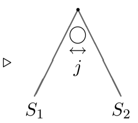
the parsing rules are very similar to the previous ones, as expected.

$$\dfrac{\begin{array}{l}\triangleright\ \textsf{stative1-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \textsf{O-S-group}(u,v,p)\end{array}}{\begin{array}{l}\triangleright\ \{+\textsf{O-S-range}(u,v), -\textsf{stative1-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \textsf{stative1}_1 \tag{7.77}$$

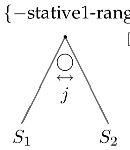
(7.78)

$$\dfrac{\begin{array}{l}\triangleright\ \textsf{stative1-range}(u,v)\\ \quad\triangleright\ \emptyset\\ \quad\triangleright\ x \in \textsf{stative1-range}(u,v) \wedge\\ \qquad\wedge\ \textsf{OJunct}(j) \in x \wedge\\ \qquad\wedge\ \textsf{Rep}(j) \in x \wedge\\ \qquad\wedge\ \textsf{Sign}(s) \in x \wedge\\ \qquad\wedge\ \textsf{Dir}(d) \in x\end{array}\quad\begin{array}{l}\triangleright\ \{\textsf{stative1-range}(u,\overleftarrow{x})\}\\ \triangleright\ \emptyset\\ \\ \triangleright\ \emptyset\\ \triangleright\ S_1[\textsf{Rep}(p_1 \Diamond c_1).\textsf{Sign}(s_1).\textsf{Dir}(d_1)]\end{array}\quad\begin{array}{l}\triangleright\ \{\textsf{stative1-range}(\overrightarrow{x},v)\}\\ \triangleright\ \emptyset\\ \vdots\pi_{\text{stative1}}\\ \triangleright\ \emptyset\\ \triangleright\ S_2[\textsf{Rep}(p_2 \Diamond c_2).\textsf{Sign}(s_2).\textsf{Dir}(d_2)]\end{array}}{\begin{array}{l}\triangleright\ \{-\textsf{stative1-range}(u,v)\}\\ \qquad[\textsf{Rep}(p_1 \Diamond p_2 \Diamond (c_1 \mathbin{\sharp_j} c_2)).\textsf{Sign}(\circlearrowright).\textsf{Dir}(\rightarrow)]\\ \qquad\qquad\bigcirc \overset{j}{\leftrightarrow} S_2\\ \qquad\triangleright\qquad\qquad S_1\end{array}}\ \textsf{stative1}_2$$

$$\frac{\begin{array}{lll} \triangleright\,\text{stative1-range}(u,v) & & \\ \triangleright\,\emptyset & & \\ \triangleright\,x \in \text{stative1-range}(u,v)\,\wedge & \triangleright\,\{\text{nom-stative-range}(\overrightarrow{u},\overleftarrow{x})\} & \triangleright\,\{\text{nom-stative-range}(\overrightarrow{x},v)\} \\ \quad\wedge\,\mathsf{OJunct}(j)\in x\,\wedge & \triangleright\,\emptyset & \triangleright\,\emptyset \\ \quad\wedge\,\mathsf{Rep}(j)\in x\,\wedge & \vdots\pi_{\text{stative}} & \vdots\pi_{\text{stative}} \\ \quad\wedge\,\mathsf{Sign}(s)\in x\,\wedge & \triangleright\,\emptyset & \triangleright\,\emptyset \\ \quad\wedge\,\mathsf{Dir}(d)\in x & \triangleright\,S_1[\mathsf{Rep}(c_1),\mathsf{Sign}(s_1),\mathsf{Dir}(d_1)] & \triangleright\,S_2[\mathsf{Rep}(c_2),\mathsf{Sign}(s_2),\mathsf{Dir}(d_2)] \end{array}}{\begin{array}{l} \triangleright\,\{-\text{stative1-range}(u,v)\} \\ \triangleright\,[\mathsf{Rep}(c_1\odot_j c_2),\mathsf{Sign}(\oplus),\mathsf{Dir}(\rightarrow)] \end{array}}\ \text{stative1}_3 \tag{7.79}$$



$$\frac{\begin{array}{lll} \triangleright\,\text{stative1-range}(u,v) & & \\ \triangleright\,\emptyset & & \\ \triangleright\,x \in \text{stative1-range}(u,v)\,\wedge & \triangleright\,\{\text{acc-stative-range}(\overrightarrow{u},\overleftarrow{x})\} & \triangleright\,\{\text{acc-stative-range}(\overrightarrow{x},v)\} \\ \quad\wedge\,\mathsf{OJunct}(j)\in x\,\wedge & \triangleright\,\emptyset & \triangleright\,\emptyset \\ \quad\wedge\,\mathsf{Rep}(j)\in x\,\wedge & \vdots\pi_{\text{stative}} & \vdots\pi_{\text{stative}} \\ \quad\wedge\,\mathsf{Sign}(s)\in x\,\wedge & \triangleright\,\emptyset & \triangleright\,\emptyset \\ \quad\wedge\,\mathsf{Dir}(d)\in x & \triangleright\,S_1[\mathsf{Rep}(c_1),\mathsf{Sign}(s_1),\mathsf{Dir}(d_1)] & \triangleright\,S_2[\mathsf{Rep}(c_2),\mathsf{Sign}(s_2),\mathsf{Dir}(d_2)] \end{array}}{\begin{array}{l} \triangleright\,\{-\text{stative1-range}(u,v)\} \\ \triangleright\,[\mathsf{Rep}(c_1\odot_{j_2} c_2),\mathsf{Sign}(\ominus),\mathsf{Dir}(\leftarrow)] \end{array}}\ \text{stative1}_4 \tag{7.80}$$



The O-S-range is parsed by extracting the preposition if present.

$$\dfrac{\begin{array}{l}\triangleright\ \text{O-S-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{O}^{+}\text{-group}(u,v,p)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{O}^{+}\text{-range}(u,v), -\text{O-S-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{O-S}_1 \tag{7.81}$$

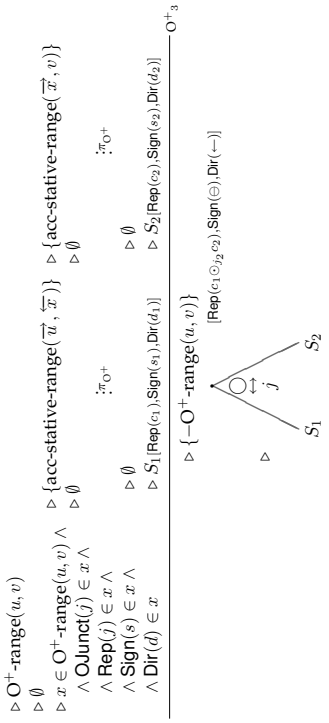
$$\dfrac{\begin{array}{ll}\triangleright\ \text{O-S-range}(u,v) & \triangleright\ \{\text{O}^{+}\text{-range}(\overrightarrow{u},v)\}\\ \triangleright\ \emptyset & \triangleright\ \emptyset\\ & \vdots\pi_{\text{O}^{+}}\\ \triangleright\ \textsf{Prep} \in u \wedge \textsf{Rep}(q) \in u \wedge & \triangleright\ \emptyset\\ \quad \wedge\ \textsf{Sign}(s) \in u \wedge \textsf{Dir}(d) \in u & \triangleright\ O_{[\textsf{Rep}(r),\textsf{Sign}(s_1),\textsf{Dir}(d_1)]}\end{array}}{\begin{array}{l}\triangleright\ \{-\text{O-S-range}(u,v)\}\\ \triangleright\ O_{[\textsf{Rep}(q \Diamond r),\textsf{Sign}(s),\textsf{Dir}(d)]}\end{array}}\ \text{O-S}_2 \tag{7.82}$$

The O^{+}-group is parsed as the stative1-group, except that there are no prepositions in the subgroups.

$$\dfrac{\begin{array}{l}\triangleright\ \text{O}^{+}\text{-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{O-group}(u,v,p)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{O-range}(u,v), -\text{O}^{+}\text{-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{O}^{+}{}_1 \tag{7.83}$$

$$\dfrac{\begin{array}{ll}\triangleright\ \text{O}^{+}\text{-range}(u,v) & \\ \triangleright\ \emptyset & \\ \triangleright\ x \in \text{O}^{+}\text{-range}(u,v) \wedge & \triangleright\ \{\text{nom-stative-range}(\overrightarrow{u},\overleftarrow{x})\} \quad \triangleright\ \{\text{nom-stative-range}(\overrightarrow{x},v)\}\\ \quad \wedge\ \textsf{OJunct}(j) \in x \wedge & \triangleright\ \emptyset\\ \quad \wedge\ \textsf{Rep}(j) \in x \wedge & \vdots\pi_{\text{O}^{+}} \qquad \vdots\pi_{\text{O}^{+}}\\ \quad \wedge\ \textsf{Sign}(s) \in x \wedge & \triangleright\ \emptyset \qquad \triangleright\ \emptyset\\ \quad \wedge\ \textsf{Dir}(d) \in x & \triangleright\ S_{1\,[\textsf{Rep}(c_1),\textsf{Sign}(s_1),\textsf{Dir}(d_1)]} \qquad \triangleright\ S_{2\,[\textsf{Rep}(c_2),\textsf{Sign}(s_2),\textsf{Dir}(d_2)]}\end{array}}{\begin{array}{l}\triangleright\ \{-\text{O}^{+}\text{-range}(u,v)\}\\ \qquad {\scriptstyle [\textsf{Rep}(c_1 \odot_j c_2),\textsf{Sign}(\oplus),\textsf{Dir}(\rightarrow)]}\\ \qquad\qquad \bigcirc \updownarrow_j \qquad S_2\\ \qquad\qquad\qquad\ \ S_1\\ \qquad\qquad \triangle\end{array}}\ \text{O}^{+}{}_2 \tag{7.84}$$

(7.85)



The O-range is parsed by the following rules, which decompose the range according to the definition.

$$\frac{\begin{array}{l}\triangleright\ \text{O-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{o-det-group}(u,v,p)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{o-det-range}(u,v), -\text{O-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{O}_1 \tag{7.86}$$

$$\frac{\begin{array}{l}\triangleright\ \text{O-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{i-group}(u,v,p)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{i-range}(u,v), -\text{O-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{O}_2 \tag{7.87}$$

$$\frac{\begin{array}{l}\triangleright\ \text{O-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{o-pron-group}(u,v,p)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{o-pron-range}(u,v), -\text{O-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{O}_3 \tag{7.88}$$

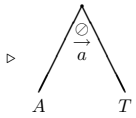
$$\frac{\begin{array}{l}\triangleright\ \text{O-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{o-emp-det-group}(u,v,p)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{o-emp-det-range}(u,v), -\text{O-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{O}_4 \tag{7.89}$$

If a determiner is present, the O-groups build a complex adtree, as already seen in section 6.2.3. The first rule parses the λ-range which forms the base of the group.

$$\frac{\begin{array}{ll}\triangleright\ \text{o-det-range}(u,v) & \triangleright\ \{\lambda\text{-range}(z,w)\}\\ \triangleright\ \emptyset & \triangleright\ \emptyset\\ \triangleright\ z \in \text{o-det-range}(u,v) \wedge & \vdots\pi_\lambda\\ \quad \wedge\ w \in \text{o-det-range}(u,v) \wedge & \triangleright\ \emptyset\\ \quad \wedge\ \lambda\text{-group}(z,w) & \triangleright\ T\end{array}}{\begin{array}{l}\triangleright\ \{+\text{o-det-1-range}(u, \overleftarrow{z}), +\text{o-det-2-range}(\overrightarrow{w}, v), -\text{o-det-range}(u,v)\}\\ \triangleright\ T\end{array}}\ \text{o-det}_1 \tag{7.90}$$

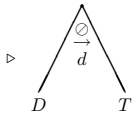
The second rule parses the modifiers (adjectives) preceding the λ-subgroup.

$$\dfrac{\begin{array}{ll} & \triangleright \{\text{Adj-range}(w, v)\} \\ & \triangleright \emptyset \\ \triangleright \text{o-det-1-range}(u, v) & \vdots \pi_{\text{Adj}} \\ \triangleright T & \\ \triangleright w \in \text{o-det-1-range}(u, v) \wedge & \triangleright \emptyset \\ \quad \wedge \text{Adj-group}(w, v, \oslash, c) & \triangleright A_{[\text{Rep}(a), \text{Sign}(\oslash), \text{Dir}(\rightarrow)]} \end{array}}{\triangleright \{+\text{o-det-1-range}(u, \overleftarrow{w}), -\text{o-det-1-range}(u, v)\}} \text{o-det}_2 \tag{7.91}$$



The third rule parses the determiner.

$$\dfrac{\begin{array}{ll} & \triangleright \{\text{det-range}(u, v)\} \\ & \triangleright \emptyset \\ \triangleright \text{o-det-1-range}(u, v) & \vdots \pi_{\text{Det}} \\ \triangleright T & \triangleright \emptyset \\ \triangleright \text{Det}(u, v, c) & \triangleright D_{[\text{Rep}(d), \text{Sign}(\oslash), \text{Dir}(\rightarrow)]} \end{array}}{\triangleright \{-\text{o-det-1-range}(u, v)\}} \text{o-det}_3 \tag{7.92}$$



The fourth and the fifth rules look on the right of the *λ*-subgroup: they parse the *-o* that ends the lexeme, along with the *-j* and the *-n*, if present.

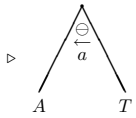
$$\dfrac{\begin{array}{l} \triangleright \text{o-det-2-range}(u, v) \wedge \neg \text{o-det-1-range}(x, y) \\ \triangleright T \\ \triangleright t \in \text{o-det-2-range}(u, v) \wedge \\ \quad \wedge \textsf{Rep}(\text{-o}) \in u \wedge \text{Agr}(\overrightarrow{u}, t, c) \end{array}}{\begin{array}{l} \triangleright \{+\text{o-det-4-range}(\overrightarrow{t}, v), -\text{o-det-2-range}(u, v)\} \\ \triangleright T_{[\text{Rep}(\text{-o} \circ c), \text{Sign}(\oplus), \text{Dir}(\rightarrow)]} \end{array}} \text{o-det}_4 \tag{7.93}$$

$$\dfrac{\begin{array}{l} \triangleright \text{o-det-2-range}(u, v) \wedge \neg \text{o-det-1-range}(x, y) \\ \triangleright T \\ \triangleright t \in \text{o-det-2-range}(u, v) \wedge \\ \quad \wedge \textsf{Rep}(\text{-o}) \in u \wedge \text{Agr}(\overrightarrow{u}, t, c) \end{array}}{\begin{array}{l} \triangleright \{+\text{o-det-4-range}(\overrightarrow{t}, v), -\text{o-det-2-range}(u, v)\} \\ \triangleright T_{[\text{Rep}(\text{-o} \circ c), \text{Sign}(\ominus), \text{Dir}(\leftarrow)]} \end{array}} \text{o-det}_5 \tag{7.94}$$

Finally, the sixth rule parses the adjectives at the end of the o-det-group.

$$\dfrac{\begin{array}{ll} & \triangleright \{\text{Adj-range}(u, w)\} \\ & \triangleright \emptyset \\ \triangleright \text{o-det-4-range}(u, v) & \\ \triangleright T_{[\text{Rep}(r), \text{Sign}(s), \text{Dir}(d)]} & \vdots \pi_{\text{Adj}} \\ \triangleright w \in \text{o-det-4-range}(u, v) \wedge & \triangleright \emptyset \\ \quad \wedge \text{Adj-group}(u, w, \ominus, c) & \triangleright A_{[\text{Rep}(a), \text{Sign}(\ominus), \text{Dir}(\leftarrow)]} \end{array}}{\triangleright \{+\text{o-det-4-range}(\overrightarrow{w}, v), -\text{o-det-4-range}(u, v)\}} \text{o-det}_6 \tag{7.95}$$

$[\text{Rep}(r), \text{Sign}(s), \text{Dir}(d)]$



The determiner is parsed according to the definition of the Det predicate.

$$\dfrac{\begin{array}{l} \triangleright \text{det-range}(u, v) \\ \triangleright \emptyset \\ \triangleright \text{DetInv}(u, v, l) \end{array}}{\begin{array}{l} \triangleright \{-\text{det-range}(u, v)\} \\ \triangleright l_{[\text{Rep}(\epsilon), \text{Sign}(\oslash), \text{Dir}(\rightarrow)]} \end{array}} \text{det}_1 \tag{7.96}$$

$$\dfrac{\begin{array}{l} \triangleright \text{det-range}(u, v) \\ \triangleright \emptyset \\ \triangleright \text{DetVar}(u, v, c, a, l) \end{array}}{\begin{array}{l} \triangleright \{-\text{det-range}(u, v)\} \\ \triangleright a_{[\text{Rep}(l \Diamond c), \text{Sign}(\oslash), \text{Dir}(\rightarrow)]} \end{array}} \text{det}_2 \tag{7.97}$$

$$\dfrac{\begin{array}{l} \triangleright \text{det-range}(u, v) \\ \triangleright \emptyset \\ \triangleright \text{DetInvCorr}(u, v, a, l) \end{array}}{\begin{array}{l} \triangleright \{-\text{det-range}(u, v)\} \\ \triangleright a_{[\text{Rep}(l), \text{Sign}(\oslash), \text{Dir}(\rightarrow)]} \end{array}} \text{det}_3 \tag{7.98}$$

$$\frac{\begin{array}{l}\triangleright\ \text{det-range}(u, v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{APron}(u, v, c, l)\end{array}}{\begin{array}{l}\triangleright\ \{-\text{det-range}(u, v)\}\\ \triangleright\ l_{[\text{Rep}(a \circ c),\text{Sign}(\oslash),\text{Dir}(\rightarrow)]}\end{array}}\ \text{det}_4 \tag{7.99}$$

$$\frac{\begin{array}{l}\triangleright\ \text{det-range}(u, v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{ANum}(u, v, l, \emptyset)\end{array}}{\begin{array}{l}\triangleright\ \{-\text{det-range}(u, v)\}\\ \triangleright\ l_{[\text{Rep}(\epsilon),\text{Sign}(\oslash),\text{Dir}(\rightarrow)]}\end{array}}\ \text{det}_5 \tag{7.100}$$

$$\frac{\begin{array}{l}\triangleright\ \text{det-range}(u, v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{ANum}(u, v, l, d) \wedge d \neq \emptyset\end{array}}{\begin{array}{l}\triangleright\ \{-\text{det-range}(u, v)\}\\ \triangleright\ \overset{\oslash}{\underset{\epsilon}{\rightarrow}}\ \begin{array}{ll} d & l \end{array}\ _{[\text{Rep}(\epsilon),\text{Sign}(\oslash),\text{Dir}(\rightarrow)]}\end{array}}\ \text{det}_6 \tag{7.101}$$

An i-range represent a verb in infinitive form, acting as a noun (Table 6.12): it is simply a λ-subgroup followed by *-i*, as expected.

$$\frac{\begin{array}{ll}\triangleright\ \text{i-range}(u, v) & \vdots\pi_\lambda\\ \triangleright\ \emptyset & \triangleright\ \emptyset\\ \triangleright\ \top & \triangleright\ T\end{array}}{\begin{array}{l}\triangleright\ \{-\text{i-range}(u, v)\}\\ \triangleright\ T_{[\text{Rep(-i)},\text{Sign}(\oplus),\text{Dir}(\rightarrow)]}\end{array}}\ \text{i}_1 \tag{7.102}$$

$$\frac{\begin{array}{ll}\triangleright\ \text{i-range}(u, v) & \vdots\pi_\lambda\\ \triangleright\ \emptyset & \triangleright\ \emptyset\\ \triangleright\ \top & \triangleright\ T\end{array}}{\begin{array}{l}\triangleright\ \{-\text{i-range}(u, v)\}\\ \triangleright\ T_{[\text{Rep(-i)},\text{Sign}(\ominus),\text{Dir}(\leftarrow)]}\end{array}}\ \text{i}_2 \tag{7.103}$$

An o-pron-group is composed by a pronoun which is parsed by the first and the second rules of this group (Table 6.11).

$$\frac{\begin{array}{l}\triangleright\ \text{o-pron-range}(u, v)\\ \triangleright\ \emptyset\\ \triangleright\ \textsf{Rep}(p) \in u\end{array}}{\begin{array}{l}\triangleright\ \{+\text{o-pron-1-range}(\overrightarrow{u}, v), -\text{o-pron-range}(u, v)\}\\ \triangleright\ p_{[\text{Rep}(\epsilon),\text{Sign}(\oplus),\text{Dir}(\rightarrow)]}\end{array}}\ \text{o-pron}_1 \tag{7.104}$$

$$\frac{\begin{array}{l}\triangleright\ \text{o-pron-range}(u, v)\\ \triangleright\ \emptyset\\ \triangleright\ \textsf{Rep}(p) \in u\end{array}}{\begin{array}{l}\triangleright\ \{+\text{o-pron-1-range}(\overrightarrow{u}, v), -\text{o-pron-range}(u, v)\}\\ \triangleright\ p_{[\text{Rep}(\epsilon),\text{Sign}(\ominus),\text{Dir}(\leftarrow)]}\end{array}}\ \text{o-pron}_2 \tag{7.105}$$

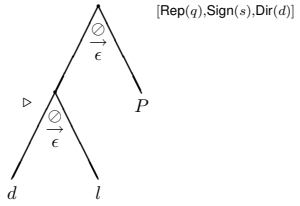
Then, the following rule parses the accusative case marker of the pronoun, if present.

$$\frac{\begin{array}{l}\triangleright\ \text{o-pron-1-range}(u, v)\\ \triangleright\ P_{[\text{Rep}(q),\text{Sign}(s),\text{Dir}(d)]}\\ \triangleright\ \textsf{Rep(-n)} \in u \wedge \textsf{Rep(-n)} \notin \overrightarrow{u}\end{array}}{\begin{array}{l}\triangleright\ \{+\text{o-pron-1-range}(\overrightarrow{u}, v), -\text{o-pron-1-range}(u, v)\}\\ \triangleright\ P_{[\text{Rep}(q \diamond \text{-n}),\text{Sign}(s),\text{Dir}(d)]}\end{array}}\ \text{o-pron}_3 \tag{7.106}$$

Then, the o-pron-group is concluded by a numerical expression. This is parsed by the following two rules, according to the absence or the presence of a modifier, respectively.

$$\frac{\begin{array}{l}\triangleright\ \text{o-pron-1-range}(u, v)\\ \triangleright\ P_{[\text{Rep}(q),\text{Sign}(s),\text{Dir}(d)]}\\ \triangleright\ \text{ANum}(u, v, l, \emptyset)\end{array}}{\begin{array}{l}\triangleright\ \{-\text{o-pron-1-range}(u, v)\}\\ \triangleright\ \overset{\oslash}{\underset{\epsilon}{\rightarrow}}\ \begin{array}{ll} l & P \end{array}\ _{[\text{Rep}(q),\text{Sign}(s),\text{Dir}(d)]}\end{array}}\ \text{o-pron}_4 \tag{7.107}$$

$$\frac{\begin{array}{l}\triangleright\ \text{o-pron-1-range}(u,v)\\ \triangleright\ P_{[\mathsf{Rep}(q),\mathsf{Sign}(s),\mathsf{Dir}(d)]}\\ \triangleright\ \text{ANum}(u,v,l,d)\wedge d\neq\emptyset\end{array}}{\triangleright\ \{-\text{o-pron-1-range}(u,v)\}}\ \text{o-pron}_5 \tag{7.108}$$

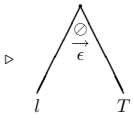


The o-emp-det-range, corresponding to the trace group, is parsed starting from the trace, represented by a void token (section 6.2.3).

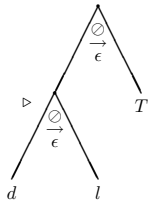
$$\frac{\begin{array}{l}\triangleright\ \text{o-emp-det-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \top\end{array}}{\begin{array}{l}\triangleright\ \{+\text{o-emp-det-1-range}(u,v),-\text{o-emp-det-range}(u,v)\}\\ \triangleright\ \square\ (\textit{Trace})\end{array}}\ \text{o-emp-det}_1 \tag{7.109}$$

Then, the numerical expression is parsed.

$$\frac{\begin{array}{l}\triangleright\ \text{o-emp-det-1-range}(u,v)\\ \triangleright\ T\\ \triangleright\ w\in\text{o-emp-det-range}(u,v)\wedge\\ \quad\wedge\ \text{ANum}(u,v,l,\emptyset)\end{array}}{\triangleright\ \{+\text{o-emp-det-1-range}(u,\overleftarrow{w}),-\text{o-emp-det-1-range}(u,v)\}}\ \text{o-emp-det}_2 \tag{7.110}$$

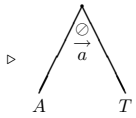


$$\frac{\begin{array}{l}\triangleright\ \text{o-emp-det-1-range}(u,v)\\ \triangleright\ T\\ \triangleright\ w\in\text{o-emp-det-range}(u,v)\wedge\\ \quad\wedge\ \text{ANum}(u,v,l,d)\wedge d\neq\emptyset\end{array}}{\begin{array}{l}\triangleright\ \{+\text{o-emp-det-1-range}(u,\overleftarrow{w}),\\ \quad-\text{o-emp-det-1-range}(u,v)\}\end{array}}\ \text{o-emp-det}_3 \tag{7.111}$$



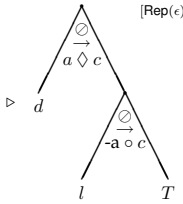
When the numerical expression has been treated, the adjectives in front of it are parsed.

$$\frac{\begin{array}{ll}\triangleright\ \text{o-emp-det-1-range}(u,v) & \triangleright\ \{\text{Adj-range}(w,v)\}\\ \triangleright\ T & \triangleright\ \emptyset\\ \triangleright\ w\in\text{o-emp-det-1-range}(u,v)\wedge & \vdots\pi_{\text{Adj}}\\ \quad\wedge\ \text{Adj-group}(w,v,\oslash,c) & \triangleright\ A_{[\mathsf{Rep}(a),\mathsf{Sign}(\oslash),\mathsf{Dir}(\rightarrow)]}\end{array}}{\triangleright\ \{+\text{o-emp-det-1-range}(u,\overleftarrow{w}),-\text{o-emp-det-1-range}(u,v)\}}\ \text{o-emp-det}_4 \tag{7.112}$$

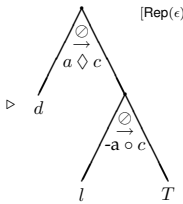


Finally, the determiner is parsed along with the pronoun.

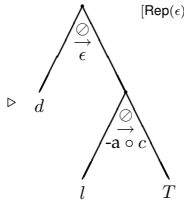
$$\frac{\begin{array}{l}\triangleright \text{o-emp-det-1-range}(u, v)\\ \triangleright T\\ \triangleright y \in \text{o-emp-det-1-range}(u, v) \wedge\\ \quad \wedge \text{APron}(y, v, c, l) \wedge \text{DetVar}(u, \overleftarrow{y}, c, a, d)\end{array}}{\begin{array}{l}\triangleright \{-\text{o-emp-det-1-range}(u, v)\}\\ \triangleright \quad d \qquad l \quad T\end{array}}\ \text{o-emp-det}_5 \tag{7.113}$$



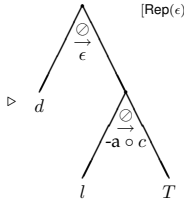
$$\frac{\begin{array}{l}\triangleright \text{o-emp-det-1-range}(u, v)\\ \triangleright T\\ \triangleright y \in \text{o-emp-det-1-range}(u, v) \wedge\\ \quad \wedge \text{APron}(y, v, c, l) \wedge \text{DetVar}(u, \overleftarrow{y}, c, a, d)\end{array}}{\begin{array}{l}\triangleright \{-\text{o-emp-det-1-range}(u, v)\}\\ \triangleright \quad d \qquad l \quad T\end{array}}\ \text{o-emp-det}_6 \tag{7.114}$$



$$\frac{\begin{array}{l}\triangleright \text{o-emp-det-1-range}(u, v)\\ \triangleright T\\ \triangleright y \in \text{o-emp-det-1-range}(u, v) \wedge\\ \quad \wedge \text{APron}(y, v, c, l) \wedge \text{DetInv}(u, \overleftarrow{y}, d)\end{array}}{\begin{array}{l}\triangleright \{-\text{o-emp-det-1-range}(u, v)\}\\ \triangleright \quad d \qquad l \quad T\end{array}}\ \text{o-emp-det}_7 \tag{7.115}$$



$$\frac{\begin{array}{l}\triangleright \text{o-emp-det-1-range}(u, v)\\ \triangleright T\\ \triangleright y \in \text{o-emp-det-1-range}(u, v) \wedge\\ \quad \wedge \text{APron}(y, v, c, l) \wedge \text{DetInv}(u, \overleftarrow{y}, d)\end{array}}{\begin{array}{l}\triangleright \{-\text{o-emp-det-1-range}(u, v)\}\\ \triangleright \quad d \qquad l \quad T\end{array}}\ \text{o-emp-det}_8 \tag{7.116}$$



7.5.5 The A-group

The A-ranges satisfy the A-group. The following rules, unfolding the definition of the A-group predicate, parse A-ranges into the corresponding adtrees.

The Adj-group may be composed or simple. Composition is analogous to the case of O-groups, so there is no need to comment the corresponding

rules.

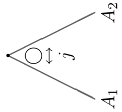
$$\frac{\begin{array}{l}\triangleright\ \text{Adj-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{Adj1-group}(u,v,s,c)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{Adj1-range}(u,v), -\text{Adj-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{Adj}_1 \tag{7.117}$$

$$\frac{\begin{array}{lll}\triangleright\ \text{Adj-range}(u,v) & & \\ \triangleright\ \emptyset & & \\ \triangleright\ x \in \text{Adj-range}(u,v) \wedge & & \\ \quad \wedge\ \text{APreJunct} \in u \wedge & & \\ \quad \wedge\ \text{Rep}(j_1) \in u \wedge & \triangleright\ \{\text{Adj1-range}(\overrightarrow{u}, \overleftarrow{x})\} & \triangleright\ \{\text{Adj1-range}(\overrightarrow{x}, v)\} \\ \quad \wedge\ \text{APostJunct}(j_1) \in x \wedge & \triangleright\ \emptyset & \triangleright\ \emptyset \\ \quad \wedge\ \text{Rep}(j_2) \in x \wedge & \vdots\pi_{\text{Adj1}} & \vdots\pi_{\text{Adj1}} \\ \quad \wedge\ \text{Sign}(s) \in x \wedge & \triangleright\ \emptyset & \triangleright\ \emptyset \\ \quad \wedge\ \text{Dir}(d) \in x & \triangleright\ A_1[\text{Rep}(c_1),\text{Sign}(s_1),\text{Dir}(d_1)] & \triangleright\ A_2[\text{Rep}(c_2),\text{Sign}(s_2),\text{Dir}(d_2)]\end{array}}{\begin{array}{l}\triangleright\ \{-\text{Adj-range}(u,v)\}\\ \triangleright\ \bullet[\text{Rep}(c_1 \odot_{j_2} c_2),\text{Sign}(\oslash),\text{Dir}(\rightarrow)] \to (A_1,\ A_2),\ \bigcirc \updownarrow j_1 \diamond j_2\end{array}}\ \text{Adj}_2 \tag{7.118}$$

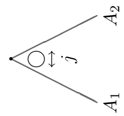
$$\frac{\begin{array}{lll}\triangleright\ \text{Adj-range}(u,v) & & \\ \triangleright\ \emptyset & & \\ \triangleright\ x \in \text{Adj-range}(u,v) \wedge & & \\ \quad \wedge\ \text{APreJunct} \in u \wedge & & \\ \quad \wedge\ \text{Rep}(j_1) \in u \wedge & \triangleright\ \{\text{Adj1-range}(\overrightarrow{u}, \overleftarrow{x})\} & \triangleright\ \{\text{Adj1-range}(\overrightarrow{x}, v)\} \\ \quad \wedge\ \text{APostJunct}(j_1) \in x \wedge & \triangleright\ \emptyset & \triangleright\ \emptyset \\ \quad \wedge\ \text{Rep}(j_2) \in x \wedge & \vdots\pi_{\text{Adj1}} & \vdots\pi_{\text{Adj1}} \\ \quad \wedge\ \text{Sign}(s) \in x \wedge & \triangleright\ \emptyset & \triangleright\ \emptyset \\ \quad \wedge\ \text{Dir}(d) \in x & \triangleright\ A_1[\text{Rep}(c_1),\text{Sign}(s_1),\text{Dir}(d_1)] & \triangleright\ A_2[\text{Rep}(c_2),\text{Sign}(s_2),\text{Dir}(d_2)]\end{array}}{\begin{array}{l}\triangleright\ \{-\text{Adj-range}(u,v)\}\\ \triangleright\ \bullet[\text{Rep}(c_1 \odot_{j_2} c_2),\text{Sign}(\ominus),\text{Dir}(\rightarrow)] \to (A_1,\ A_2),\ \bigcirc \updownarrow j_1 \diamond j_2\end{array}}\ \text{Adj}_3 \tag{7.119}$$

$$\frac{\begin{array}{l}\triangleright\ \text{Adj1-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{A-group}(u,v,s,c)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{A-range}(u,v), -\text{Adj1-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{Adj1}_1 \tag{7.120}$$

$$\frac{\begin{array}{l}\triangleright\ \text{Adj1-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ x \in \text{Adj1-range}(u,v) \wedge\\ \quad \wedge\ \mathsf{AJunct} \in x \wedge\\ \quad \wedge\ \mathsf{Rep}(j) \in x \wedge\\ \quad \wedge\ \mathsf{Sign}(s) \in x \wedge\\ \quad \wedge\ \mathsf{Dir}(d) \in x\end{array} \quad \begin{array}{l}\triangleright\ \{\text{Adj1-range}(u, \overleftarrow{x})\}\\ \triangleright\ \emptyset\\ \quad \vdots \pi_{\text{Adj1}}\\ \triangleright\ \emptyset\\ \triangleright\ A_1[\mathsf{Rep}(c_1),\mathsf{Sign}(s_1),\mathsf{Dir}(d_1)]\end{array} \quad \begin{array}{l}\triangleright\ \{\text{Adj1-range}(\overrightarrow{x}, v)\}\\ \triangleright\ \emptyset\\ \quad \vdots \pi_{\text{Adj1}}\\ \triangleright\ \emptyset\\ \triangleright\ A_2[\mathsf{Rep}(c_2),\mathsf{Sign}(s_2),\mathsf{Dir}(d_2)]\end{array}}{\begin{array}{l}\triangleright\ \{-\text{Adj1-range}(u,v)\}\\ \triangleright\ [\mathsf{Rep}(c_1 \odot_j c_2),\mathsf{Sign}(\oslash),\mathsf{Dir}(\rightarrow)]\end{array}}\ \text{Adj1}_2 \tag{7.121}$$



$$\frac{\begin{array}{l}\triangleright\ \text{Adj1-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ x \in \text{Adj1-range}(u,v) \wedge\\ \quad \wedge\ \mathsf{AJunct} \in x \wedge\\ \quad \wedge\ \mathsf{Rep}(j) \in x \wedge\\ \quad \wedge\ \mathsf{Sign}(s) \in x \wedge\\ \quad \wedge\ \mathsf{Dir}(d) \in x\end{array} \quad \begin{array}{l}\triangleright\ \{\text{Adj1-range}(u, \overleftarrow{x})\}\\ \triangleright\ \emptyset\\ \quad \vdots \pi_{\text{Adj1}}\\ \triangleright\ \emptyset\\ \triangleright\ A_1[\mathsf{Rep}(c_1),\mathsf{Sign}(s_1),\mathsf{Dir}(d_1)]\end{array} \quad \begin{array}{l}\triangleright\ \{\text{Adj1-range}(\overrightarrow{x}, v)\}\\ \triangleright\ \emptyset\\ \quad \vdots \pi_{\text{Adj1}}\\ \triangleright\ \emptyset\\ \triangleright\ A_2[\mathsf{Rep}(c_2),\mathsf{Sign}(s_2),\mathsf{Dir}(d_2)]\end{array}}{\begin{array}{l}\triangleright\ \{-\text{Adj1-range}(u,v)\}\\ \triangleright\ [\mathsf{Rep}(c_1 \odot_j c_2),\mathsf{Sign}(\ominus),\mathsf{Dir}(\leftarrow)]\end{array}}\ \text{Adj1}_3 \tag{7.122}$$



An A-range may represent an A1-group (Table 6.17), an A2-group (Table 6.18) a prepositional clause or an AJoined group, i.e., a phrase used as a clause.

$$\frac{\begin{array}{l}\triangleright\ \text{A-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{A1-group}(u,v,c)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{A1-range}(u,v), -\text{A-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{A}_1 \tag{7.123}$$

$$\frac{\begin{array}{l}\triangleright\ \text{A-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{A2-group}(u,v,c)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{A2-range}(u,v), -\text{A-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{A}_2 \tag{7.124}$$

$$\frac{\begin{array}{l}\triangleright\ \text{A-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{circ-group}(u,v)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{circ-range}(u,v), -\text{A-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{A}_3 \tag{7.125}$$

$$\frac{\begin{array}{l}\triangleright\ \text{A-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{AJoined}(u,v)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{AJoined-range}(u,v), -\text{A-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{A}_3 \tag{7.126}$$

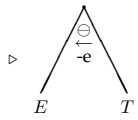
An A1-group is a λ-group which forms its kernel with some additional structure. The kernel is parsed by the following rules.

$$\frac{\begin{array}{ll}\triangleright\ \text{A1-range}(u,v) & \triangleright\ \{\lambda\text{-range}(w,\overleftarrow{t}\,)\}\\ \triangleright\ \emptyset & \triangleright\ \emptyset\\ \triangleright\ w \in \text{A1-range}(u,v) \wedge & \vdots\pi_\lambda\\ \quad \wedge\, t \in \text{A1-range}(u,v) \wedge & \triangleright\ \emptyset\\ \quad \wedge\, \textsf{Rep}(\text{-a}) \in t \wedge \lambda\text{-group}(w,\overleftarrow{t}\,) \wedge \text{Agr}(\overrightarrow{t}\,,v,c) & \triangleright\ T\end{array}}{\begin{array}{l}\triangleright\ \{+\text{A1-range}(u,\overleftarrow{w}), -\text{A1-range}(u,v)\}\\ \triangleright\ T_{[\textsf{Rep}(\text{-a}\circ c),\textsf{Sign}(\oslash),\textsf{Dir}(\rightarrow)]}\end{array}}\ \text{A1}_1 \tag{7.127}$$

$$\frac{\begin{array}{ll}\triangleright\ \text{A1-range}(u,v) & \triangleright\ \{\lambda\text{-range}(w,\overleftarrow{t}\,)\}\\ \triangleright\ \emptyset & \triangleright\ \emptyset\\ \triangleright\ w \in \text{A1-range}(u,v) \wedge & \vdots\pi_\lambda\\ \quad \wedge\, t \in \text{A1-range}(u,v) \wedge & \triangleright\ \emptyset\\ \quad \wedge\, \textsf{Rep}(\text{-a}) \in t \wedge \lambda\text{-group}(w,\overleftarrow{t}\,) \wedge \text{Agr}(\overrightarrow{t}\,,v,c) & \triangleright\ T\end{array}}{\begin{array}{l}\triangleright\ \{+\text{A1-range}(u,\overleftarrow{w}), -\text{A1-range}(u,v)\}\\ \triangleright\ T_{[\textsf{Rep}(\text{-a}\circ c),\textsf{Sign}(\ominus),\textsf{Dir}(\leftarrow)]}\end{array}}\ \text{A1}_2 \tag{7.128}$$

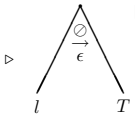
The first piece of structure on the right of the kernel is a sequence of e-groups. The following rule takes care of parsing them.

$$\frac{\begin{array}{ll}\triangleright\ \text{A1-range}(u,v) & \triangleright\ \{\text{E-range}(w,\overleftarrow{v}\,)\}\\ \triangleright\ T_{[\textsf{Rep}(a),\textsf{Sign}(s),\textsf{Dir}(d)]} & \triangleright\ \emptyset\\ \triangleright\ w \in \text{A1-range}(u,v) \wedge & \vdots\pi_{\text{E}}\\ \quad \wedge\, \text{e-group}(w,v) & \triangleright\ \emptyset\\ & \triangleright\ E\end{array}}{\triangleright\ \{+\text{A1-range}(u,\overleftarrow{w}), -\text{A1-range}(u,v)\}}\ \text{A1}_3 \tag{7.129}$$

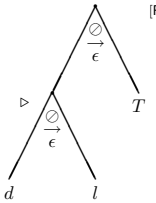


Then an ANum may be found. As usual, there are two rules depending on the presence or absence of a modifier.

$$\frac{\begin{array}{l}\triangleright\ \text{A1-range}(u,v)\\ \triangleright\ T_{[\textsf{Rep}(a),\textsf{Sign}(s),\textsf{Dir}(d)]}\\ \triangleright\ x \in \text{A1-range}(u,v) \wedge\\ \quad \wedge\, \text{ANum}(x,v,l,\emptyset)\end{array}}{\triangleright\ \{+\text{A1-1-range}(u,\overleftarrow{x}), -\text{A1-range}(u,v)\}}\ \text{A1}_4 \tag{7.130}$$

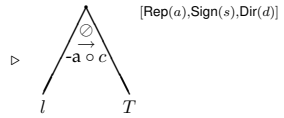


$$\dfrac{\begin{array}{l}\triangleright\ \text{A1-range}(u,v)\\ \triangleright\ T^{[\text{Rep}(a),\text{Sign}(s),\text{Dir}(d)]}\\ \triangleright\ x \in \text{A1-range}(u,v) \wedge \\ \quad \wedge\ \text{ANum}(x,v,l,d) \wedge d \neq \emptyset\end{array}}{\triangleright\ \{+\text{A1-1-range}(u, \overleftarrow{x}), -\text{A1-range}(u,v)\}}\ \text{A1}_5 \tag{7.131}$$

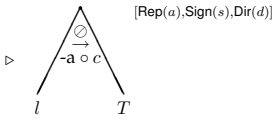


Finally, an APron may be found. It gets parsed by the following rules.

$$\dfrac{\begin{array}{l}\triangleright\ \text{A1-range}(u,v)\\ \triangleright\ T^{[\text{Rep}(a),\text{Sign}(s),\text{Dir}(d)]}\\ \triangleright\ \text{APron}(u,v,c,l)\end{array}}{\triangleright\ \{-\text{A1-range}(u,v)\}}\ \text{A1}_6 \tag{7.132}$$



$$\dfrac{\begin{array}{l}\triangleright\ \text{A1-1-range}(u,v)\\ \triangleright\ T^{[\text{Rep}(a),\text{Sign}(s),\text{Dir}(d)]}\\ \triangleright\ \text{APron}(u,v,c,l)\end{array}}{\triangleright\ \{-\text{A1-1-range}(u,v)\}}\ \text{A1}_7 \tag{7.133}$$

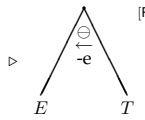


Since A2-groups are symmetrical to A-group, there is no need to comment their parsing rules. They have an analogous format as the corresponding A1 rules.

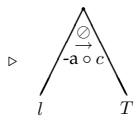
$$\dfrac{\begin{array}{ll}\triangleright\ \text{A2-range}(u,v) & \triangleright\ \{\lambda\text{-range}(w, \overleftarrow{t}\,)\}\\ \triangleright\ \emptyset & \triangleright\ \emptyset\\ \triangleright\ w \in \text{A2-range}(u,v) \wedge & \vdots\pi_\lambda\\ \quad \wedge\ t \in \text{A2-range}(u,v) \wedge & \triangleright\ \emptyset\\ \quad \wedge\ \text{Rep(-a)} \in t \wedge \lambda\text{-group}(w, \overleftarrow{t}\,) \wedge \text{Agr}(\overrightarrow{t}\,, v, c) & \triangleright\ T\end{array}}{\begin{array}{l}\triangleright\ \{+\text{A2-range}(u, \overleftarrow{w}), -\text{A2-range}(u,v)\}\\ \triangleright\ T^{[\text{Rep(-a}\circ c),\text{Sign}(\oslash),\text{Dir}(\rightarrow)]}\end{array}}\ \text{A2}_1 \tag{7.134}$$

$$\dfrac{\begin{array}{ll}\triangleright\ \text{A2-range}(u,v) & \triangleright\ \{\lambda\text{-range}(w, \overleftarrow{t}\,)\}\\ \triangleright\ \emptyset & \triangleright\ \emptyset\\ \triangleright\ w \in \text{A2-range}(u,v) \wedge & \vdots\pi_\lambda\\ \quad \wedge\ t \in \text{A2-range}(u,v) \wedge & \triangleright\ \emptyset\\ \quad \wedge\ \text{Rep(-a)} \in t \wedge \lambda\text{-group}(w, \overleftarrow{t}\,) \wedge \text{Agr}(\overrightarrow{t}\,, v, c) & \triangleright\ T\end{array}}{\begin{array}{l}\triangleright\ \{+\text{A2-range}(u, \overleftarrow{w}), -\text{A2-range}(u,v)\}\\ \triangleright\ T^{[\text{Rep(-a}\circ c),\text{Sign}(\ominus),\text{Dir}(\leftarrow)]}\end{array}}\ \text{A2}_2 \tag{7.135}$$

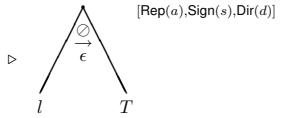
$$\dfrac{\begin{array}{ll}\triangleright\ \text{A2-range}(u,v) & \triangleright\ \{\text{E-range}(w, \overleftarrow{v}\,)\}\\ \triangleright\ T^{[\text{Rep}(a),\text{Sign}(s),\text{Dir}(d)]} & \triangleright\ \emptyset\\ \triangleright\ w \in \text{A2-range}(u,v) \wedge & \vdots\pi_\text{E}\\ \quad \wedge\ \text{e-group}(w,v) & \triangleright\ \emptyset\\ & \triangleright\ E\end{array}}{\triangleright\ \{+\text{A2-range}(u, \overleftarrow{w}), -\text{A2-range}(u,v)\}}\ \text{A2}_3 \tag{7.136}$$



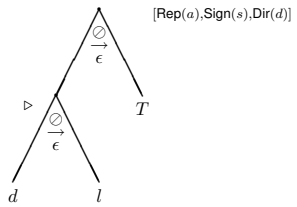
$$\frac{\begin{array}{l}\triangleright\ \text{A2-range}(u,v)\\ \triangleright\ T^{[\text{Rep}(a),\text{Sign}(s),\text{Dir}(d)]}\\ \triangleright\ x \in \text{A2-range}(u,v) \wedge \\ \quad \wedge\ \text{APron}(x,v,c,l)\end{array}}{\triangleright\ \{+\text{A2-1-range}(u,\overleftarrow{x}), -\text{A2-range}(u,v)\}}\ \text{A2}_4 \tag{7.137}$$

$\triangleright$ 

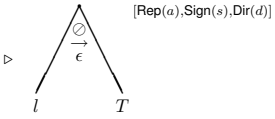
$$\frac{\begin{array}{l}\triangleright\ \text{A2-range}(u,v)\\ \triangleright\ T^{[\text{Rep}(a),\text{Sign}(s),\text{Dir}(d)]}\\ \triangleright\ \text{ANum}(u,v,l,\emptyset)\end{array}}{\triangleright\ \{-\text{A2-range}(u,v)\}}\ \text{A2}_5 \tag{7.138}$$

$\triangleright$ 

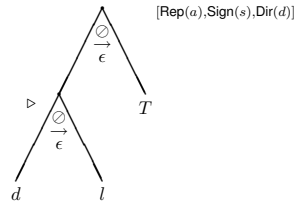
$$\frac{\begin{array}{l}\triangleright\ \text{A2-range}(u,v)\\ \triangleright\ T^{[\text{Rep}(a),\text{Sign}(s),\text{Dir}(d)]}\\ \triangleright\ \text{ANum}(u,v,l,d) \wedge d \neq \emptyset\end{array}}{\triangleright\ \{-\text{A2-range}(u,v)\}}\ \text{A2}_6 \tag{7.139}$$



$$\frac{\begin{array}{l}\triangleright\ \text{A2-1-range}(u,v)\\ \triangleright\ T^{[\text{Rep}(a),\text{Sign}(s),\text{Dir}(d)]}\\ \triangleright\ \text{ANum}(u,v,l,\emptyset)\end{array}}{\triangleright\ \{-\text{A2-1-range}(u,v)\}}\ \text{A2}_7 \tag{7.140}$$

$\triangleright$ 

$$\frac{\begin{array}{l}\triangleright\ \text{A2-1-range}(u,v)\\ \triangleright\ T^{[\text{Rep}(a),\text{Sign}(s),\text{Dir}(d)]}\\ \triangleright\ \text{ANum}(u,v,l,d) \wedge d \neq \emptyset\end{array}}{\triangleright\ \{-\text{A2-1-range}(u,v)\}}\ \text{A2}_6 \tag{7.141}$$



$$\frac{\begin{array}{ll}\triangleright\ \text{AJoined-range}(u,v) & \vdots\pi_{\text{phrase}}\\ \triangleright\ \emptyset & \triangleright\ \emptyset\\ \triangleright\ \text{Rep}(a) \in u \wedge \text{Sign}(s) \wedge \text{Dir}(d) & \triangleright\ P\end{array}}{\begin{array}{l}\triangleright\ \{-\text{AJoined-range}(u,v)\}\\ \triangleright\ P^{[\text{Rep}(a),\text{Sign}(s),\text{Dir}(d)]}\end{array}}\ \text{AJoined} \tag{7.142}$$

$\triangleright\ \{\text{phrase-range}(\overrightarrow{u},v)\}$
$\triangleright\ \emptyset$

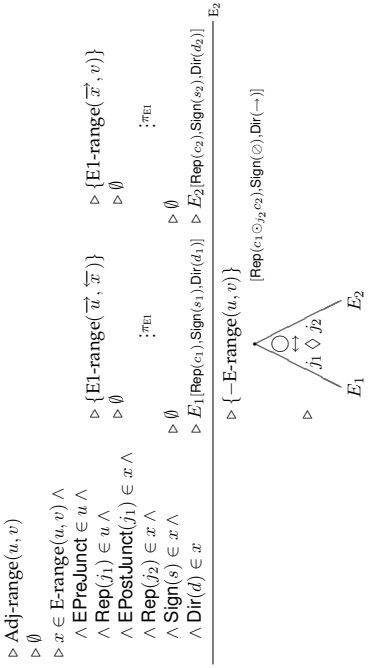
7.5.6 The E-group

The circumstantial group contains the rules to parse the circumstantials – and the clauses applied to verbs too – appearing in the phrase. They are never part of the valence-based verbal tree and they never modify a stative group.

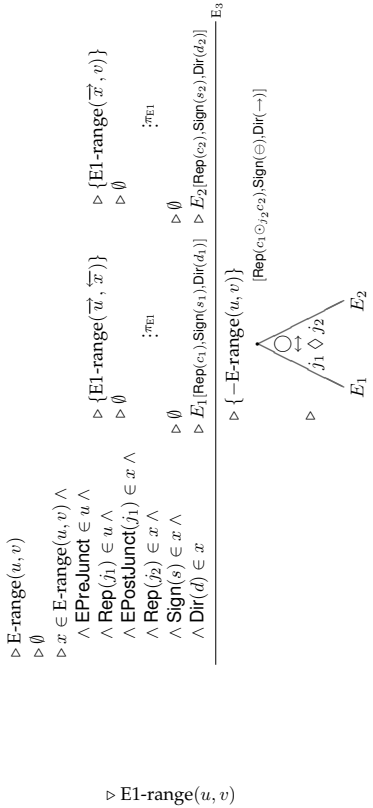
An E-group may be a composed by means of junctions. The rules are completely analogous to the ones for the O-group and the A-group. They are listed below without any further comment.

$$\frac{\begin{array}{l}\triangleright\ \text{E-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{E1-group}(u,v)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{E1-range}(u,v), -\text{E-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{E}_1 \tag{7.143}$$

(7.144)

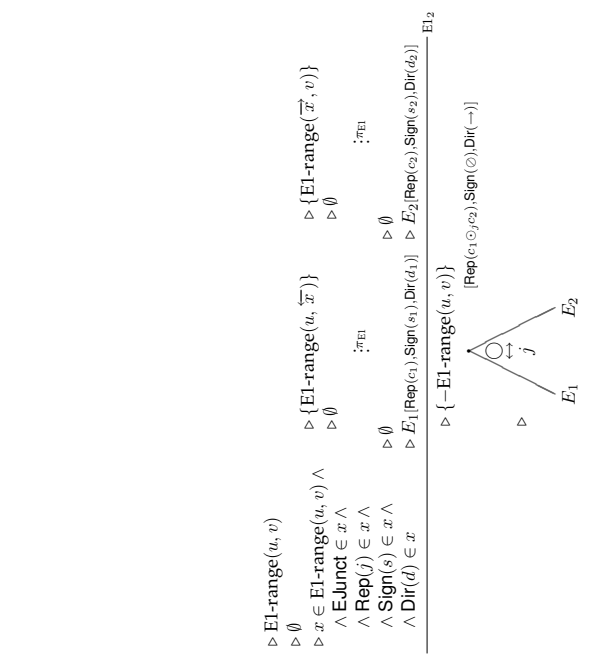


(7.145)

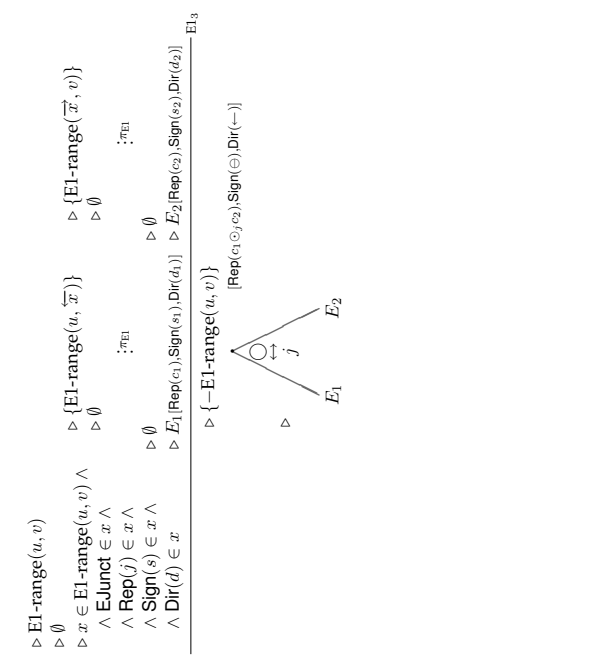


$$\frac{\begin{array}{l}\triangleright\ \text{E1-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{E-S-group}(u,v)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{E-S-range}(u,v), -\text{E1-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{E1}_1 \tag{7.146}$$

(7.147)



(7.148)



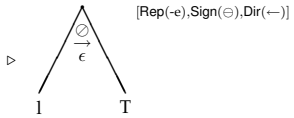
A simple E-group, E-S-group for short, is a circumstantial or a prepositional clause. Since it admits modifiers, it requires some rules to parse the various possibilities.

(7.149)

$$\dfrac{\begin{array}{ll}\triangleright\ \text{E-S-range}(u,v) & \\ \triangleright\ A & \\ \triangleright\ \textsf{Rep(-e)} \in v \wedge \textsf{EDETINV} \notin u & \end{array} \qquad \begin{array}{l}\triangleright\ \{\lambda\text{-range}(u, \overleftarrow{v})\} \\ \triangleright\ \emptyset \\ \vdots\pi_\lambda \\ \triangleright\ \emptyset \\ \triangleright\ T\end{array}}{\begin{array}{l}\triangleright\ \{-\text{E-S-range}(u,v)\} \\ \triangleright\ T_{[\textsf{Rep(-e)},\textsf{Sign}(\ominus),\textsf{Dir}(\leftarrow)]}\end{array}}\ \text{E-S}_1$$

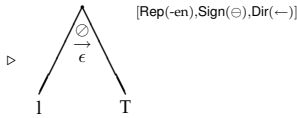
$$\dfrac{\begin{array}{ll}\triangleright \text{E-S-range}(u,v) & \\ \triangleright A & \\ \triangleright \textsf{Rep}(\text{-e}) \in \overleftarrow{v} \wedge \textsf{Rep}(\text{-n}) \in v \wedge \textsf{EDETINV} \notin u & \end{array} \quad \begin{array}{l} \triangleright \{\lambda\text{-range}(u, \overleftarrow{v})\} \\ \triangleright \emptyset \\ \vdots \pi_\lambda \\ \triangleright \emptyset \\ \triangleright T \end{array}}{\begin{array}{l} \triangleright \{-\text{E-S-range}(u,v)\} \\ \triangleright T_{[\textsf{Rep}(\text{-en}),\textsf{Sign}(\ominus),\textsf{Dir}(\leftarrow)]} \end{array}}\ \text{E-S}_2 \tag{7.150}$$

$$\dfrac{\begin{array}{ll}\triangleright \text{E-S-range}(u,v) & \\ \triangleright A & \\ \triangleright \textsf{Rep}(\text{-e}) \in v \wedge & \\ \quad \wedge \textsf{EDETINV} \in u \wedge \textsf{Rep}(l) \in u & \end{array} \quad \begin{array}{l} \triangleright \{\lambda\text{-range}(u, \overleftarrow{v})\} \\ \triangleright \emptyset \\ \vdots \pi_\lambda \\ \triangleright \emptyset \\ \triangleright T \end{array}}{\triangleright \{-\text{E-S-range}(u,v)\}}\ \text{E-S}_3$$

$\triangleright$ 

(7.151)

$$\dfrac{\begin{array}{ll}\triangleright \text{E-S-range}(u,v) & \\ \triangleright A & \\ \triangleright \textsf{Rep}(\text{-e}) \in \overleftarrow{v} \wedge \textsf{Rep}(\text{-n}) \in v & \\ \quad \wedge \textsf{EDETINV} \in u \wedge \textsf{Rep}(l) \in u & \end{array} \quad \begin{array}{l} \triangleright \{\lambda\text{-range}(u, \overleftarrow{v})\} \\ \triangleright \emptyset \\ \vdots \pi_\lambda \\ \triangleright \emptyset \\ \triangleright T \end{array}}{\triangleright \{-\text{E-S-range}(u,v)\}}\ \text{E-S}_4$$

$\triangleright$ 

(7.152)

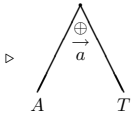
If the simple E-range is a prepositional clause, the corresponding range

to parse is circ-range.

$$\dfrac{\begin{array}{l}\triangleright \text{E-S-range}(u,v) \\ \triangleright A \\ \triangleright \text{circ-group}(u,v)\end{array}}{\begin{array}{l}\triangleright \{-\text{E-S-range}(u,v), +\text{circ-range}(u,v)\} \\ \triangleright \emptyset\end{array}}\ \text{E-S}_5 \tag{7.153}$$

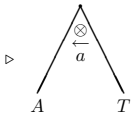
Prepositional clauses may be attached to the tree on the left or on the right, according to the sign of the adposition. Moreover, it should be noticed that the following rules are deputed to parse a sequence of clauses and circumstantials: this apparently counterintuitive format for the parsing rules becomes very useful in the parsing of phrase groups indeed.

$$\dfrac{\begin{array}{ll}\triangleright \text{circ-range}(x,y) & \\ \triangleright A & \\ \triangleright u \in \text{circ-range}(x,y) & \end{array} \quad \begin{array}{l} \triangleright \{\text{prep-stative-range}(u,y)\} \\ \triangleright \emptyset \\ \vdots \pi_{\text{prep-stative}} \\ \triangleright \emptyset \\ \triangleright T_{[\textsf{Rep}(a),\textsf{Sign}(\oplus),\textsf{Dir}(\rightarrow)]} \end{array}}{\begin{array}{l} \triangleright \{+\text{circ-range}(x, \overleftarrow{u}), -\text{circ-range}(x,y)\} \\ \triangleright \end{array}}\ \text{Circ}_\oplus$$



(7.154)

$$\dfrac{\begin{array}{ll}\triangleright \text{circ-range}(x,y) & \\ \triangleright A & \\ \triangleright u \in \text{circ-range}(x,y) & \end{array} \quad \begin{array}{l} \triangleright \{\text{prep-stative-range}(u,y)\} \\ \triangleright \emptyset \\ \vdots \pi_{\text{prep-stative}} \\ \triangleright \emptyset \\ \triangleright T_{[\textsf{Rep}(a),\textsf{Sign}(\otimes),\textsf{Dir}(\leftarrow)]} \end{array}}{\begin{array}{l} \triangleright \{+\text{circ-range}(x, \overleftarrow{u}), -\text{circ-range}(x,y)\} \\ \triangleright \end{array}}\ \text{Circ}_\otimes$$



(7.155)

$$\dfrac{\begin{array}{ll}\triangleright\ \text{circ-range}(x,y) & \\ \triangleright\ A & \\ \triangleright\ u \in \text{circ-range}(x,y) & \end{array}\qquad \begin{array}{l}\triangleright\ \{\text{prep-stative-range}(u,y)\} \\ \triangleright\ \emptyset \\ \vdots \pi_{\text{prep-stative}} \\ \triangleright\ \emptyset \\ \triangleright\ T_{[\mathsf{Rep}(a),\mathsf{Sign}(\ominus),\mathsf{Dir}(\leftarrow)]}\end{array}}{\begin{array}{l}\triangleright\ \{+\text{circ-range}(x,\overleftarrow{u}), -\text{circ-range}(x,y)\} \\ \triangleright\ \overset{\ominus}{\underset{\leftarrow}{a}}(T,A)\end{array}}\ \text{Circ}_{\ominus} \tag{7.156}$$

$$\dfrac{\begin{array}{ll}\triangleright\ \text{circ-range}(x,y) & \\ \triangleright\ A & \\ \triangleright\ u \in \text{circ-range}(x,y) & \end{array}\qquad \begin{array}{l}\triangleright\ \{\text{prep-stative-range}(u,y)\} \\ \triangleright\ \emptyset \\ \vdots \pi_{\text{prep-stative}} \\ \triangleright\ \emptyset \\ \triangleright\ T_{[\mathsf{Rep}(a),\mathsf{Sign}(\oslash),\mathsf{Dir}(\rightarrow)]}\end{array}}{\begin{array}{l}\triangleright\ \{+\text{circ-range}(x,\overleftarrow{u}), -\text{circ-range}(x,y)\} \\ \triangleright\ \overset{\oslash}{\underset{\rightarrow}{a}}(T,A)\end{array}}\ \text{Circ}_{\oslash} \tag{7.157}$$

$$\dfrac{\begin{array}{ll}\triangleright\ \text{circ-range}(u,v) & \\ \triangleright\ A & \\ \triangleright\ x \in \text{circ-range}(u,v) & \end{array}\qquad \begin{array}{l}\triangleright\ \{\text{E-range}(x,v)\} \\ \triangleright\ \emptyset \\ \vdots \pi_{\text{E}} \\ \triangleright\ \emptyset \\ \triangleright\ T_{[\mathsf{Rep}(a),\mathsf{sign}(\ominus),\mathsf{dir}(\leftarrow)]}\end{array}}{\begin{array}{l}\triangleright\ \{+\text{circ-range}(u,\overleftarrow{x}), \\ \quad -\text{circ-range}(u,v)\} \\ \triangleright\ \overset{\ominus}{\underset{\leftarrow}{a}}(T,A)\end{array}}\ \text{Circ}_{\text{E}} \tag{7.158}$$

$$\dfrac{\begin{array}{l}\triangleright\ \text{prep-stative-range}(x,y) \\ \triangleright\ \emptyset \\ \triangleright\ \text{prep-nom-stative-group}(x,y,p) \vee \\ \quad \vee\ \text{prep-acc-stative-group}(x,y,p)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{stative-range}(x,y), -\text{prep-stative-range}(x,y)\} \\ \triangleright\ \emptyset\end{array}}\ \text{prep-stative} \tag{7.159}$$

Since e-groups are a subcase of E-S-groups, which, in turn, are a special case of E-groups, there is no need to have specific rules to parse them. In fact, after the e-group has been checked by the associated predicate, it can be parsed by the rules governing the E-range.

At last, the EJoined group acts exactly as the AJoined group.

$$\dfrac{\begin{array}{ll}\triangleright\ \text{EJoined-range}(u,v) & \\ \triangleright\ \emptyset & \\ \triangleright\ \mathsf{Rep}(a) \in u \wedge \mathsf{Sign}(s) \wedge \mathsf{Dir}(d) & \end{array}\qquad \begin{array}{l}\triangleright\ \{\text{phrase-range}(\overrightarrow{u},v)\} \\ \triangleright\ \emptyset \\ \vdots \pi_{\text{phrase}} \\ \triangleright\ \emptyset \\ \triangleright\ P\end{array}}{\begin{array}{l}\triangleright\ \{-\text{EJoined-range}(u,v)\} \\ \triangleright\ P_{[\mathsf{Rep}(a),\mathsf{Sign}(s),\mathsf{Dir}(d)]}\end{array}}\ \text{AJoined} \tag{7.160}$$

7.5.7 The phrase group

This group of rules is deputed to parse the structure of a phrase. They act by looking at the valence of the verb and then by constructing the appropriate tree.

$$\dfrac{\begin{array}{ll}\triangleright\ \text{phrase-range}(u,v) & \\ \triangleright\ \emptyset & \\ \triangleright\ x \in \text{phrase-range}(u,v) \wedge & \\ \quad \wedge\ y \in \text{phrase-range}(u,v) \wedge & \\ \quad \wedge\ \text{I-group}(x,y,0,d,c) & \end{array}\qquad \begin{array}{l}\triangleright\ \{\text{I-range}(x,y)\} \\ \triangleright\ \emptyset \\ \vdots \pi_{\lambda} \\ \triangleright\ \emptyset \\ \triangleright\ V\end{array}}{\begin{array}{l}\triangleright\ \{+\text{circ-range}(u,\overleftarrow{x}), +\text{circ-range}(\overrightarrow{y},v), -\text{phrase-range}(u,v)\} \\ \triangleright\ \overset{\oplus}{\underset{\rightarrow}{d}}(\square, V\ (\textit{Verb}))\ (\textit{Phrase})\end{array}}\ \text{Verb}_0 \tag{7.161}$$

The whole adtree is labelled as $Phrase$, while the verbal subtree is labelled as $Verb$. These labels are used to determine the actants after the whole text has been successfully parsed.[24]

To simplify the notation we introduce the following abbreviations:

$$\begin{array}{ccccc} \vdots & \equiv & \vdots & \equiv & \begin{array}{l} \triangleright\, \{\text{I-range}(u,v)\} \\ \triangleright\, \emptyset \\ \vdots\pi_\lambda \\ \triangleright\, \emptyset \\ \triangleright\, V \end{array} \\ \mathcal{D}_{\text{verb}} & & \mathcal{D}_{\text{verb}}(u,v,V) & & \end{array} \tag{7.162}$$

and

$$\begin{array}{ccccc} \vdots & \equiv & \vdots & \equiv & \begin{array}{l} \triangleright\, \{\text{subj-range}(u,v)\} \\ \triangleright\, \emptyset \\ \vdots\pi_{\text{subj}} \\ \triangleright\, \{\text{circ-range}(u,s_1), \\ \quad \text{circ-range}(s_2,v)\} \\ \triangleright\, S_{[\text{Sign}(\oplus),\text{Dir}(\rightarrow),\text{Rep}(s)]} \end{array} \\ \mathcal{D}_{\text{subj}} & & \mathcal{D}_{\text{subj}}(u,v,s_1,s_2,s,S) & & \end{array} \tag{7.163}$$

where the variables may be hidden when they are evident from the context.

The Verb_1 rule is very similar to the corresponding zerovalent counterpart, as expected. The main difference is that the subject is no more a void token, but it has to be found in the subj-range by the derivation π_{subj}. In the subj-range there may be other parts that cannot be but circumstantials, hence the conclusion of the subderivation.

$$\frac{\begin{array}{l} \triangleright\, \text{phrase-range}(u,v) \\ \triangleright\, \emptyset \\ \triangleright\, x \in \text{phrase-range}(u,v) \wedge \\ \quad \wedge\, y \in \text{phrase-range}(u,v) \wedge \\ \quad \wedge\, \text{I-group}(x,y,1,d,c) \end{array} \qquad \begin{array}{cc} \vdots & \vdots \\ \mathcal{D}_{\text{verb}}(x,y,V) & \mathcal{D}_{\text{subj}}(u,\overleftarrow{x},s_1,s_2,s,S) \end{array}}{\begin{array}{l} \triangleright\, \{+\text{circ-range}(u,s_1), +\text{circ-range}(s_2,\overleftarrow{x}), \\ \quad +\text{circ-range}(\overrightarrow{y},v), -\text{phrase-range}(u,v)\} \\ \triangleright \end{array} \begin{array}{c} (\textit{Phrase}) \\ \overset{\oplus}{\rightarrow} \\ s \Diamond d \\ S \qquad V\ (\textit{Verb}) \end{array}}\ \text{Verb}_1 \tag{7.164}$$

The Verb_1 rule assumes to find the subject on the left of the verb. Instead, when the subject appears on the right, mirroring has taken place.

The following rule takes care of this situation.[25]

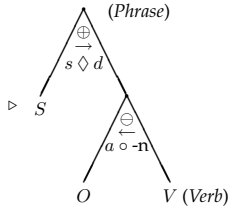
$$\frac{\begin{array}{l} \triangleright\, \text{phrase-range}(u,v) \\ \triangleright\, \emptyset \\ \triangleright\, x \in \text{phrase-range}(u,v) \wedge \\ \quad \wedge\, y \in \text{phrase-range}(u,v) \wedge \\ \quad \wedge\, \text{I-group}(x,y,1,d,c) \end{array} \qquad \begin{array}{cc} \vdots & \vdots \\ \mathcal{D}_{\text{verb}}(x,y,V) & \mathcal{D}_{\text{subj}}(\overrightarrow{y},v,s_1,s_2,s,S) \end{array}}{\begin{array}{l} \triangleright\, \{+\text{circ-range}(u,\overleftarrow{x}), +\text{circ-range}(\overrightarrow{y},s_1), \\ \quad +\text{circ-range}(s_2,v), -\text{phrase-range}(u,v)\} \\ \triangleright \end{array} \begin{array}{c} (\textit{Phrase}) \\ \overset{\ominus}{\leftarrow} \\ d \Diamond s \\ V\ (\textit{Verb}) \qquad S \end{array}}\ \text{Verb}_{1M} \tag{7.165}$$

The following abbreviation is useful to compact notation:

$$\begin{array}{ccccc} \vdots & \equiv & \vdots & \equiv & \begin{array}{l} \triangleright\, \{\text{obj-range}(u,v)\} \\ \triangleright\, \emptyset \\ \vdots\pi_{\text{obj}} \\ \triangleright\, \{\text{circ-range}(u,o_1), \\ \quad \text{circ-range}(o_2,v)\} \\ \triangleright\, O_{[\text{Rep}(ao\text{-n}),\text{Dir}(\leftarrow),\text{Sign}(\ominus)]} \end{array} \\ \mathcal{D}_{\text{obj}} & & \mathcal{D}_{\text{obj}}(u,v,o_1,o_2,a,S) & & \end{array} \tag{7.166}$$

The Verb$_2$ rule is an extension of the monovalence rule.

$$\dfrac{\begin{array}{l}\triangleright\ \text{phrase-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ x\in\text{phrase-range}(u,v)\ \wedge\\ \quad\wedge\, y\in\text{phrase-range}(u,v)\ \wedge \quad \vdots \quad \vdots \quad \vdots\\ \quad\wedge\,\text{I-group}(x,y,2,d,c) \qquad \mathcal{D}_{\text{verb}}\ \ \mathcal{D}_{\text{subj}}\ \ \mathcal{D}_{\text{obj}}(\overrightarrow{y},v,o_1,o_2,a,S)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{circ-range}(u,s_1),+\text{circ-range}(s_2,\overleftarrow{x}),+\text{circ-range}(\overrightarrow{y},o_1),\\ \quad +\text{circ-range}(o_2,v),-\text{phrase-range}(u,v)\}\end{array}}\ \text{Verb}_2$$



(7.167)

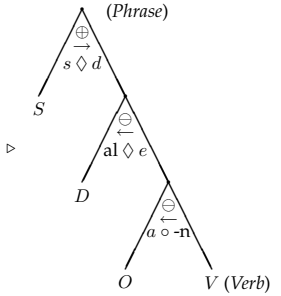
As usual, an abbreviation for the dative is used:

$$\begin{array}{ccccc}\vdots & \equiv & \vdots & \equiv & \triangleright\ \{\text{dat-range}(u,v)\}\\ \mathcal{D}_{\text{dat}} & & \mathcal{D}_{\text{dat}}(u,v,d_1,d_2,d,D) & & \triangleright\ \emptyset\\ & & & & \vdots_{\pi_{\text{dat}}}\\ & & & & \triangleright\ \{\text{circ-range}(u,d_1),\\ & & & & \quad\text{circ-range}(d_1,v)\}\\ & & & & \triangleright\ D_{[\text{Rep}(\text{al}\lozenge d),\text{Sign}(\ominus),\text{Dir}(\leftarrow)]}\end{array}$$

(7.168)

The Verb$_3$ rule is an extension of the rules for bivalent verbs.

$$\dfrac{\begin{array}{l}\triangleright\ \text{phrase-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ x\in\text{phrase-range}(u,v)\ \wedge\\ \quad\wedge\, y\in\text{phrase-range}(u,v)\ \wedge \ \vdots \quad \vdots \quad \vdots \quad \vdots\\ \quad\wedge\,\text{I-group}(x,y,3,d,c) \qquad \mathcal{D}_{\text{verb}}\ \ \mathcal{D}_{\text{subj}}\ \ \mathcal{D}_{\text{obj}}\ \ \mathcal{D}_{\text{dat}}(o_2,v,\dots)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{circ-range}(u,s_1),+\text{circ-range}(s_2,\overleftarrow{x}),+\text{circ-range}(\overrightarrow{y},o_1),\\ \quad +\text{circ-range}(o_2,d_1),+\text{circ-range}(d_2,v),-\text{phrase-range}(u,v)\}\end{array}}\ \text{Verb}_3$$



(7.169)

7.5.8 The subject-object-dative group

These rules are deputed to parse the substructures directly related to a verb, i.e., the subject group, the object group and the dative group. Evidently, these are used to sligthly simplify the notation in the phrase group or, at least, to allow the reader to intuitively interpret the meaning of the rules.

As the name suggests, the Subject rule parses the syntactic subject of the current phrase. Its action is to choose a stative group in the nominative case inside the subject range and to parse it in the subderivation. The final tree contains the subject and the state has a pair of circumstantial ranges on the left and on the right of the subject. The tree is labelled to identify the position of the subject in the final tree so that actants can be correctly

derived.

$$\dfrac{\begin{array}{ll}\triangleright\ \text{subj-range}(x,y) & \triangleright\ \{\text{stative-range}(u,v)\}\\ \triangleright\ \emptyset & \triangleright\ \emptyset\\ \triangleright\ u \in \text{subj-range}(x,y) \wedge & \vdots \pi_{\text{stative}}\\ \quad \wedge\, v \in \text{subj-range}(x,y) \wedge & \triangleright\ \emptyset\\ \quad \wedge\, u \overset{=}{\longleftarrow} v \wedge \text{nom-stative-group}(u,v) & \triangleright\ S[\text{Rep}(s),\text{Sign}(\oplus),\text{Dir}(\rightarrow)]\end{array}}{\begin{array}{l}\triangleright\ \{+\text{circ-range}(x,\overleftarrow{u}),+\text{circ-range}(\overrightarrow{v},y),-\text{subj-range}(x,y)\}\\ \triangleright\ S[\text{Rep}(s),\text{Sign}(\oplus),\text{Dir}(\rightarrow)]\ (\textit{Subject})\end{array}}\ \text{Subject}$$

(7.170)

The object is parsed by the Object rule. It closely resembles the Subject rule except that the object must be in the accusative case. The tree is labelled to identify the position of the subject in the final tree so that actants can be correctly derived.

$$\dfrac{\begin{array}{ll}\triangleright\ \text{obj-range}(x,y) & \triangleright\ \{\text{stative-range}(u,v)\}\\ \triangleright\ \emptyset & \triangleright\ \emptyset\\ \triangleright\ u \in \text{obj-range}(x,y) \wedge & \vdots \pi_{\text{stative}}\\ \quad \wedge\, v \in \text{obj-range}(x,y) \wedge & \triangleright\ \emptyset\\ \quad \wedge\, u \overset{=}{\longleftarrow} v \wedge \text{acc-stative-group}(u,v) & \triangleright\ O[\text{Rep}(l),\text{Dir}(\leftarrow),\text{Sign}(\ominus)]\end{array}}{\begin{array}{l}\triangleright\ \{+\text{circ-range}(x,\overleftarrow{u}),+\text{circ-range}(\overrightarrow{v},y),-\text{obj-range}(x,y)\}\\ \triangleright\ O[\text{Rep}(l),\text{Dir}(\leftarrow),\text{Sign}(\ominus)]\ (\textit{Object})\end{array}}\ \text{Object}$$

(7.171)

The dative of a trivalent verb has the form 'al $\circ\, N'$, where N is a stative group. The Dative rule checks that there is a properly formed group in the dative range, and parses it. The resulting tree is labelled with the *al* preposition.

$$\dfrac{\begin{array}{ll}\triangleright\ \text{dat-range}(x,y) & \\ \triangleright\ \emptyset & \triangleright\ \{\text{stative-range}(u,v)\}\\ \triangleright\ u \in \text{dat-range}(x,y) \wedge & \triangleright\ \emptyset\\ \quad \wedge\, v \in \text{dat-range}(x,y) \wedge & \vdots \pi_{\text{stative}}\\ \quad \wedge\, u \longleftarrow v \wedge & \triangleright\ \emptyset\\ \quad \wedge\, \text{prep-nom-stative-group}(u,v,\text{al} \Diamond e) & \triangleright\ D[\text{Rep}(\text{al}\Diamond e),\text{Dir}(\leftarrow),\text{Sign}(\ominus)]\end{array}}{\begin{array}{l}\triangleright\ \{+\text{circ-range}(x,\overleftarrow{u}),+\text{circ-range}(\overrightarrow{v},y),-\text{dat-range}(x,y)\}\\ \triangleright\ D[\text{Rep}(\text{al}\Diamond e),\text{Dir}(\leftarrow),\text{Sign}(\ominus)]\ (\textit{Dative})\end{array}}\ \text{Dative}$$

(7.172)

7.5.9 The sentence group

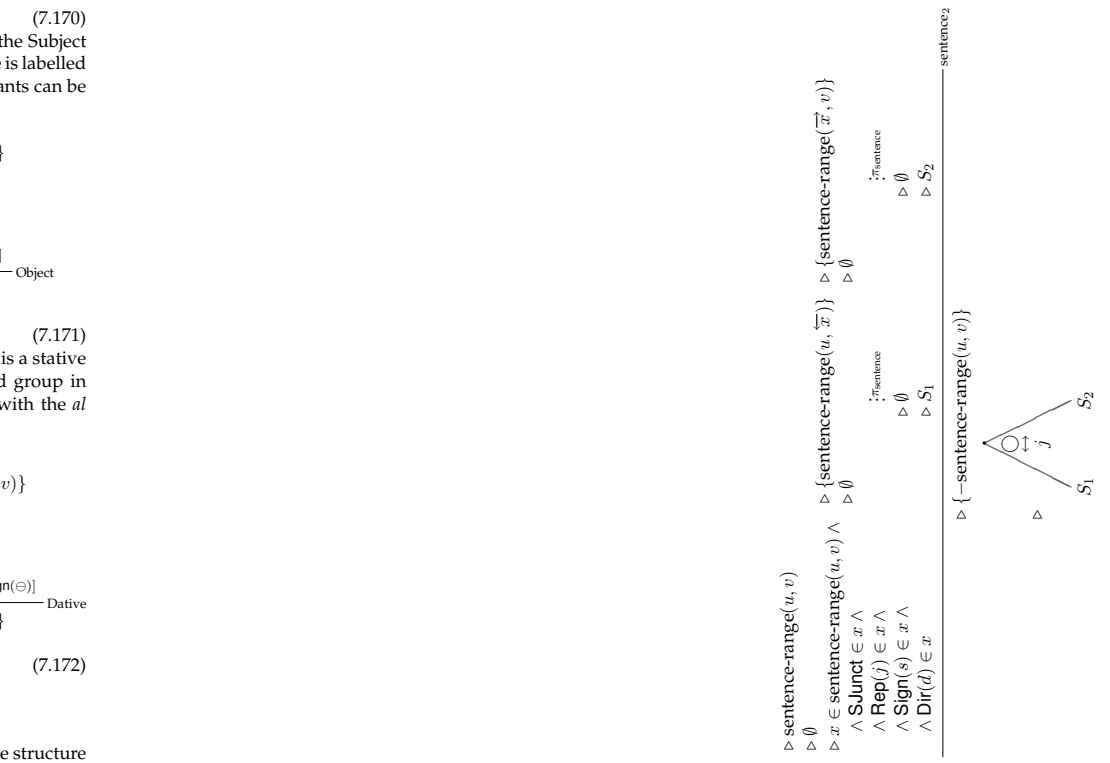
This group of rules wants to divide a sentence into a phrases. The structure of a sentence is composed by phrases in a completely analogous way to the

structure or E-groups, A-groups and O-groups (Table 6.20).

$$\dfrac{\begin{array}{l}\triangleright\ \text{sentence-range}(u,v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{phrase-group}\end{array}}{\begin{array}{l}\triangleright\ \{+\text{phrase-range}(u,v),-\text{sentence-range}(u,v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{sentence}_1$$

(7.173)

(7.174)

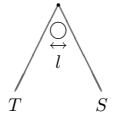


7.5.10 The text group

This group of rules divides the text into sentences, which are parsed independently. They are joined in a sequence by means of the sentence delimiters (section 5.4).

$$\frac{\begin{array}{l}\triangleright\ \text{text-range}(u, v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{sentence-group}(u, v)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{sentence-range}(u, v), -\text{text-range}(u, v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{text}_1 \tag{7.175}$$

$$\frac{\begin{array}{l}\triangleright\ \text{text-range}(u, v)\\ \triangleright\ T\\ \triangleright\ x \in \text{text-range}(u, v) \wedge\\ \quad \wedge\ \text{sentence-group}(u, \overleftarrow{x}) \wedge\\ \quad \wedge\ \mathsf{Rep}(l) \in x \wedge \mathsf{SentenceDelimiter} \in x \wedge\\ \quad \wedge\ \mathsf{Sign}(s) \in x \wedge \mathsf{Dir}(d) \in x\end{array} \qquad \begin{array}{l}\triangleright\ \{\text{sentence-range}(u, \overleftarrow{x})\}\\ \triangleright\ \emptyset\\ \vdots \pi_{\text{sentence}}\\ \triangleright\ \emptyset\\ \triangleright\ S\end{array}}{\triangleright\ \{+\text{text-range}(\overrightarrow{x}, v), -\text{text-range}(u, v)\}}\ \text{text}_2$$

$\triangleright$ 

(7.176)

$$\frac{\begin{array}{l}\triangleright\ \text{init}(u, v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{text-group}(u, v)\end{array}}{\begin{array}{l}\triangleright\ \{+\text{text.range}(u, v), -\text{init}(u, v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{init}_2 \tag{7.177}$$

$$\frac{\begin{array}{l}\triangleright\ \text{init}(u, v)\\ \triangleright\ \emptyset\\ \triangleright\ \text{text-group}(u, \overleftarrow{v}) \wedge\\ \quad \wedge\ \mathsf{Rep}(l) \in x \wedge \mathsf{SentenceDelimiter} \in x \wedge\\ \quad \wedge\ \mathsf{Sign}(s) \in x \wedge \mathsf{Dir}(d) \in x\end{array}}{\begin{array}{l}\triangleright\ \{+\text{text.range}(u, v), -\text{init}(u, v)\}\\ \triangleright\ \emptyset\end{array}}\ \text{init}_2 \tag{7.178}$$

$$\frac{\begin{array}{l}\triangleright\ \text{init}(u, v)\\ \triangleright\ \emptyset\\ \triangleright\ u \rightarrowtail v\end{array}}{\begin{array}{l}\triangleright\ \{-\text{init}(u, v)\}\\ \triangleright\ \square\end{array}}\ \text{init}_3 \tag{7.179}$$

List of Tags

Tag	Argument description	Group
AdjCorr		O
AHJunct		A
AJunct		A
APostJunct(x)	an adtree	A
APreJunct		A
Atom(x)	I-atom, O-atom, A-atom, E-atom	λ
Dir(x)	$\leftarrow$, $\rightarrow$	Parsing rules
EDetEnv		E
EHJunct		E
EJunct		E
EPostJunct(x)	an adtree	E
EPreJunct		E
Fundamental		λ
G-prefix		λ
G-suffix		λ
I		I
I-prefix		λ
I-suffix		λ
InvAdjCorr		O
Invar		O
Num		A
NumAtom(x)	numerical agreement (Table 5.41)	A
NumDet		A
O-prefix		λ
O-suffix		λ
OJunct		O
OPostJunct(x)	an adtree	O
OPreJunct		O
Prep		O
Pron		O,A
Rep(x)	an adtree	λ, I, O, A, E
Sign(x)	$\oplus$, $\ominus$, $\otimes$, $\oslash$	Parsing rules
SentenceDelimiter		text
SJunct		sentence
Val(x)	0, 1, 2, 3	λ
VarAdjCorr		O

Notes

[1] Sometimes more adtrees are admitted after the parsing of the input string. If so, the final selection will be made on the basis of semantics, eventually lead by pragmatic information such as actants. In other words, semantic ambiguity should not be solved on a syntactic level. This is not covered by the formal model.

[2] For Von Neumann's architecture see, e.g., von Neumann [1981] and Goldstine and von Neumann [1963].

[3] For non-determinism in algorithms, see Savage [1997] and Cormen et al. [1990]. For the notion of backtracking, see Wirth [1986].

[4] For a discussion about this point see, e.g., Partington [1997] or Aït-Kaci [1991].

[5] It can be adapted to any NL in principle, but I have chosen Esperanto for the reasons already explained in chapter 4.

[6] See Furlan and Lanzarone [1988] for an introduction and Sterling and Shapiro [1994] for advanced aspects about Prolog. The most used logical frameworks at the time of writing (November 2008) are: LCF [Paulson, 1987], Isabelle [Paulson, 1990] and ALF [Coquand et al., 1994]. I want to mention apartly the Grammatical Framework (GF), which is a logical framework that different formalisms (even CCG, [Steedman and Baldridge, 2007]) specialised for multilingual grammar writing – see at least Ranta [2004] and Ranta [2007] for technicalities.

[7] This list is partially based on Wennergren [2005, §1.2].

[8] At the limit, the tag Rep(m) can be empty ($m = 0$) in phrases like *se mastro as fore*. In other words, even fundamental finals can be lexicalised: in the case just offered, the 'lexeme' *as* inherits the moduli of the lexeme *est* (to be), in particular the valence.

[9] Thus, if $a \notin T$ then $a \leftarrow\!\!\!\!\shortmid\ b$ and $b \mapsto\!\!\!\!\shortmid\ a$ are both false.

[10] Please note that the epsilon (ϵ) is also used to mark empty hooks in adtrees.

[11] In a way, λ stands for a 'complex lexeme' – λ is the Greek {l}, the first letter in *lexeme*. λ also stands for 'abstraction' in mathematical logic, therefore it seems to be highly appropriate.

[12] In formulae, the string 'nom-' indicates 'nominative', while 'acc-' indicates 'accusative'.

[13] I use the terms of traditional grammar if they are clear enough and know, as in the case of 'dative'.

[14] This became clear to me at first actually writing the rules: the formalism let understand better NL mechanisms, while a better NL understanding permit to expand the formalism accordingly [Gualandri, 2008]. I prefer not to use the term 'complement' as it is really abused. For example, in systemic grammar, it has a pragmatic value, indicating adjectives or statives which refer either to the first or the second valence arguments anaphorically or cataphorically.

[15] As a technical detail, it should be noted that, if $r = \emptyset$ or $r = \epsilon$, then $q \lozenge r = q$. So, empty adpositions act as neutral elements when calculating a composed adposition.

[16] I use the term 'junction' as a hyperonym of prepositions, junctives, etc..

[17] From now on, silently, I will treat the other junctions in the same way.

[18] A clever dictionay may be implemented by, e.g., a RDF or an OWL ontology. In this case, the deductive ability of the ontology is used to implement the complex finite state automaton deputed to recognise numerical objects.

[19] As a last note, this justifies why I consider that theoretical linguistics should always test its models formally and computationally, as I claim in the introduction.

[20] From a technical point of view, there is no need of mirroring shown in Figure 2.34 was not implemented yet, essentially for timing.

[21] In a preliminary version of this dissertation, a derivation tree was planned to be put here: theoretically speaking, it could be useful for the over-formal reader once to see how the algorithm parses a string and builds an adtree. However, a monophrasal sentence with five groups – not too simple, not too complex – would easily occupy 40 A4-pages of this dissertation. Let us consider as an example the O-group *la libr-o*, 'the book', which is only

one group with only three tokens. Let us assume that we know in advance that it is a stative group, let us say, a syntactic subject. Let us also assume that we do not need non-determinism, i.e., that, thanks to an oracle, the machine always chooses the right direction in the derivation tree. Even with a very simple group (not a phrase!), and two *strong* assumptions, the parsing algorithm is too precise in order to be transcribed on paper. In fact, the algorithm starts computing the first clause – as it is the simplest case – of the following predicates: 9, 10, 11, 12, 14, 15. Only when arrived at predicate 15, it *starts* analysing the inner structure of the group itself: what computed until here is of no interest for a human reader, even an over-formal one.

[22] In practice, the Struct_2 is never used. But, in perspective, when dealing with a language less regular than Esperanto, it seems to be unavoidable, so I preferred to leave it in the system.

[23] The adposition in the hook has $\bigcirc$ as sign and $\leftrightarrow$ as direction. This is a shortcut to say that the sign is s, the sign of the adposition and the direction is d, the natural direction of the adposition. The direction influences the positions of the S_1 and the S_2 trees: if the direction is $\rightarrow$, then S_1 is the left child while S_2 lies on the right, as shown in the rule; if the direction is $\leftarrow$, then S_2 is on the right and S_1 is on the left, in reverse order with respect to the tree in the conclusion of the rule. From now on, I silently assume the above interpretation of $\bigcirc$ and $\leftrightarrow$.

[24] I hope that this use of fonts helps the reader to understand what I am saying.

[25] From now on, the variant rules required to cope with mirroring are omitted. The reader may find it instructing to write them by himself.

Chapter 8

A machine translation example

In chapter 4 I have presented the theoretical paradigms of MT and their current implementations as corpora-driven system. In this small chapter I present the system architecture through an example.

8.1 The system architecture

The MT system proposed here is a rule-based, transfer system, where the input entry / source language $\mathcal{L}_i$ is always Esperanto, while the output / target language $\mathcal{L}_o$ can be any NL. MT is performed in two steps:

1. metataxis;
2. substitution.

A zero step is postulated to answer to the question of efficiency (computationally) and of the rule/list fallacy (linguistically). The zero step is a query in the TM, where correspondent adpositional patterns are stored: this means that the differences between $\mathcal{T}_o$ and $\mathcal{T}_1$ are registered, based on adtypes, grammar characters (and, eventually, in the future, actants).

If even a single difference is registered, the MT system performs the metataxis. The term 'metataxis' is a calque by Schubert [1987] of the Tesnerian term *metataxe*. In adgrams, the term **metataxis** indicates the layer of the transformation; more specifically, from an adtree $\mathcal{T}_i$ belonging to the input language to an adtree $\mathcal{T}_o$ belonging to the output language $\mathcal{L}_o$. At the end of the metataxis, the result is a sort of pivot language. In other words, the lexicon, i.e., the tags put on the leaves, including rep(d), still pertain $\mathcal{L}_i$, i.e., the QNL Esperanto, while the adtree $\mathcal{T}_o$, pertains the NL in input.[1]

It there is no difference between $\mathcal{T}_o$ and $\mathcal{T}_1$, or, when metataxis is finished, the second step takes place. I call the second step **substitution**, i.e., the substitution of the morphemes from the dictionary of the input language ($\mathcal{D}_i$) to the dictionary of the output language ($\mathcal{D}_o$). In the translation

game scenario, these two steps are performed twice, i.e., for the source and target language, until the user is satisfied, and then the results are put appropriately in the TM. This is the end.[2]

As an extreme rare case, the MT system can be successful in finding the *whole* text in input in its TMs: e.g., if an English Esperantophone write in input William Auld's book *La Mastro de l' Ringoj*, he or she will see on the screen J. R. R. Tolkien's *The Lord of the Rings*, if this masterpiece is already in the dictionary, starting from Tolkien's poem: *Tri ringoj por la elfoj sub la hela ĉiel', Sep por la gnomoj en salonoj el ŝton'*. (Figure 8.1).

Figure 8.1: The machine translation of *The Lord of the Rings* (specimen).

8.2 A translation game example

Suppose that Alice is an English native and an Esperantophone, who accepted to use the system. She wants to write a letter to a Chinese friend of hers, Li. They do not have any (Q)NL in common.

- (46-eo.) La biblioteko kie mi ofte studas literaturon estas tre fore.
- (46-en.) The library where I often study literature is far away.

Alice types example 46 into the system, in Esperanto, and reads on the screen the English translation, that she consider satisfying. What happened in the machine when Alice finished typing her text in Esperanto, and has clicked the button 'test' so to control her Esperanto through the English translation? In fact, no metataxis was performed, as the two adtrees are structurally identical: only substitution was needed – – see Figures 8.2 and 8.3. What Li is expected to read is the following sentence.

- (46-zh.) wǒ jīng chāng xuéxí wén xué de túshū guǎn lí zhér hěn yuǎn.

In this case metataxis comes into action, as Esperanto and Chinese are structurally different – Figure 8.4. Most notably, what in Esperanto and English is a special adpositional pattern, i.e., the circumstantial correlative

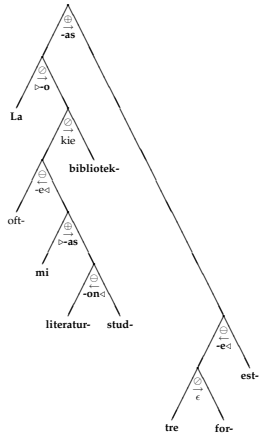


Figure 8.2: The MT sentence in Esperanto (example 46).

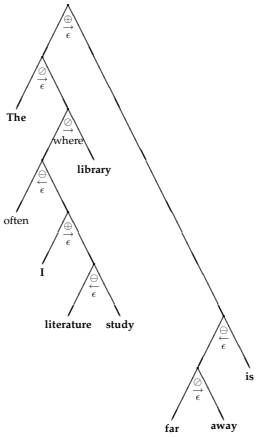


Figure 8.3: The MT sentence in English (example 46).

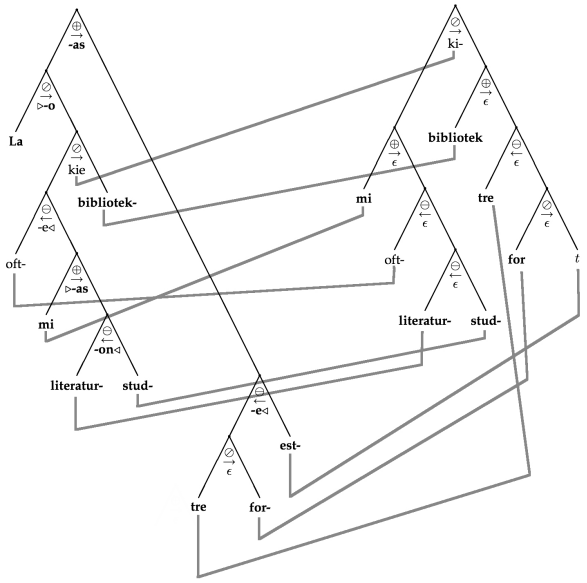


Figure 8.4: Metataxis: from Esperanto to "Chinese-anto" (example 46)

kie, in Chinese is simply a junctive. As the sentence level is completed, the MT system pass to the phrase level, starting from the main phrase. In Chinese there is no root *est-*, nor example 46 shows explicit hooks: the adtree structure is given by collocation, and the lexical base is void (□) syntactically. Then, the secondary phrase is parsed analogously. Note that *stud-* and *literatur-* are almost preserved, while the article *la* is simply omitted. The result is a "Chinese-anto" sentence.

As metataxis is completed, the system takes the second step, i.e., it performs substitution (Figure 8.5).[3] The reader is invited to check the shape of

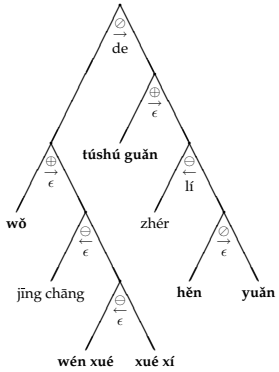


Figure 8.5: The MT sentence in Chinese (example 46).

the "Chinese-anto" adtree and the final Chinese adtree: it is clear that the second and last step of MT is again substitution.

8.3 Beyond the translation game

The MT architecture is based on Esperanto, because of the translation game, but it can be generalised. My aim is to implement adpositional grammars as an abstract, general linguistic model. In fact, the notions of adpositional space, adpositional type, grammar character, transferer and transferee are completely general: the translation game can can be generalised with any couple of NLs, there $\mathcal{L}_e$ is the entry language and $\mathcal{L}_t$ is the target language.

$$A > \mathcal{L}_e \Rightarrow B < \mathcal{L}_t$$

Nonetheless, such an approach would be much more expensive in every sense: e.g., for English, adpositionally there would be a lot of transferers

and transferees which can be applied only to some lexemes for phonological or historical reasons, as the small fragment presented in chapter 3 has attested. Most probably, an additional set of rules should be written to pass from a regularised quasi-natural English to a subset of natural English.

What I plan to do in the future is to refactor the rules, in a more abstract and polished way, so to generate all and only the cross-linguistical structures, while the intricacies of the Esperanto adpositional grammar – chapter 5 has shown that there are, even if far less than in NLs – being part of the lexicon.[4] Perhaps, an additional layer would be needed, so not to complicate too much the linguistic rules in the dictionary, in case of NLs. In other words, it is worth to retain the QNL as a key concept, so that the model would produce before something like *mouse-s run-ed quick-y* in quasi-English.[5] The resulting final architecture would be something like this:

- the **grammar level**: a general and cross-linguistical grammar (composed by the metataxis level, the substitution level, and maybe by the pragmatic layer);
- the **QNL-level**: a number of dictionaries for the various QNL (quasi-English, quasi-Italian, etc.);
- the **NL-level**: hard-coded rules for handling exceptions, and perhaps idioms and lexicalisations too.

The implementation of the pragmatic layer – i.e., how to compute actants and their relation with valence – was postponed after this dissertation, as it is important that that layer does not interfere with the metataxis level for linguistic reasons. In other words, before implementing actants, it makes sense to have a set of rules cross-linguistically valid but totally indipendent by the actual (Q)NL used.

The final architecture can be considered belonging to the interlingua paradigm, but with an important difference: normally interlinguas computed semantics, here what is computed is the core of adpositional grammars.

Notes

[1] Adpositional patterns are similar to the concept of example as posed by Nagao [1984], with the exception that for him an example is made of words, while for me an adpositional pattern is made of pure structure, with dedicated morphemes as vehicles.

[2] Under a theoretical point of view, this architecture can be described as a transfer system concerning metataxis and as a direct system concerning substitution.

[3] Note, that, because of Hanyu Pinyin, Chinese reveals lexicalisations, e.g., *túshú guǎn* (often). This is only a technical fact.

[4] For instance, see section 6.3 about the parsing of affixes, for intricacies in the Esperanto grammar.

[5] Troyanskii, the forgotten pioneer of MT using Esperanto, would have written the same example in the following way: *mouse-J run-IS quick-E* [Hutchins and Lovtskii, 2000]. His would-be writing is equivalent.

Chapter 9

Conclusions

This dissertation presented adpositional grammars, a multilingual grammar formalism. The first part of the dissertation has shown that adpositional grammars are cognitively well-founded on the very general dichotomy trajector/landmark, and linguistically robust, as provided by the examples in natural languages typologically very distant, like English, Turkish. Complex phenomena as mirroring are also covered in the model (chapters 1, 2). The translation game (chapter 4) has offered a representation of machine translation that is unique and it is free from prejudice.

The second part of this dissertation has shown that adpositional grammars are computable, as they can be implemented with a strong formalism such as a logical framework. For the purposes of this dissertation, the implementation in a simple, regular and powerful quasi-natural language as Esperanto has shown the linguistic viability of the model (chapter 5). I claim that the model presented here can parse and represent more than 95% of the standard register of contemporary Esperanto, testable through the existing monolingual and bilingual web corpora.

The third part has shown the implementation as an instance of the Esperanto grammar in a strong formalism (chapters 6, 7). In fact, the formal model let to find the number of Esperanto language: five. The main goal, i.e., to prove the consistency and robustness of the formal model both under a computationally and linguistically points of view: the implementation of Esperanto in exactly 56 formulas has shown it clearly. Furthermore, the grammar can be generalised for many natural language, and how to do this was explained in various points of the third part.

Current limits

There are some linguistic and computational limits in the current version of the adpositional grammar model. Linguistically, there is no resolution of anaphoras, actants, and non-neutral orders, such as the ones caused by

mirroring (however, the appropriate strutures are already in the model, or I know where to put them, if needed). On a more technical level, the adpositional space does not take into account neutral adpositions as posed by Pennacchietti in his original work, as it is not clear how to represent them formally and therefore computationally.

From a computational point of view, the most evident limit is pretty obvious: the actual implementation is 'only' in Latex – by the way, me and Marco Benini plan to release the package for representing adtrees on CTAN as soon as possible (by now, it lacks only the necessary documentation). Furthermore, the implementation proposed is not necessarily efficient, also because of the rule/list fallacy, i.e., the fact that this formalism does not take into account any storage of rules, except for numbers and question correlatives. However, efficiency was never a goal here.

Finally, the model takes into account the adpositional tree structure of Esperanto, along with valence – modifiers included – but there is still no treatment of actants, i.e., the pragmatic layer. These evident limits are essentially caused by timing. In the next months, I plan to fill this gap starting from the example presented in the second part of this dissertation and giving a consistent fragment, eventually implementing the algorithm for actants, as described in chapter 2, analogously to what presented in chapter 7. The actants are important for machine translation purposes, as briefly shown in chapter 8.

Further research directions

After a polishing of the rules presented in chapters 7 and 8, two results should be obtained: first, a generalisation of the grammar, second, a even more robust formalism. The way seems to be to abstract lexemes, pronouns, etc., in general, pre-linguistic groups – e.g., λ-group for lexeme, π-group for pronouns, δ-groups for determiners, etc. That means that adpositional grammars are an object more general than chomskian grammars, i.e., where terminals are no more terminals. Let me use a metaphor: adpositional grammars are a topology, where lexemes are like points, and hence the grammar of a (Q)NL is a concrete topology, where the metataxis is a formal topology. As a corollary, this dissertation shows that linguistics can be considered an instance of contemporary mathematics. This fact has some concrete consequences in the current model. First, the entry language is transformed to an abstract representation of cognitive linguistic knowledge, and hence in the translation game any (quasi-)natural language can be source or target. In other words, the grammar will not be linked necessarily to Esperanto. The final goal is to cut off all non-deterministic points – as what was used is the satisfiability algorithm.

Also the nature of translation changes dramaticaly. In fact, from a more

general and abstract point of view, the correct family of adtrees of a given sentence can be considered the place where every point is zero, i.e., a 'variety', in algebraic geometry. Consequentially, the translation from natural language to natural language is the study of every possible manipolation of equations, without touching the lexicon. Eventually, this has lead us (me and Marco Benini) think that translation is solvable in P-SPACE.

Acknowledgements

I wish to acknowledge some people. Roberto, the owner of Pottery Cafe, via Solferino, Milan, who provided great vegan and non-vegan cakes to me and Marco Benini during our many brainstorming meetings. Emanuele Gualandri, Albertina Bonini and Caterina Schweizer, who gave me a house in Ginostra and the tranquillity to work in one of the beautiful places I know in the world (in the meantime, they also gave a big help to me and my wife with our children). Umberto Mazza, who helped me with the psychological aspects related to the role of being a PhD. The album *Accelerate* by R.E.M. was my soundtrack during these difficult and exciting months. Ettore Brocca, Andrea Gervasini, Domenico Marini, Massimiliano Pepe, once students and now friends, were encouraging me keeping discretely and gently in touch, even during the most difficult moments: thanks, guys! It is impossible to list here all other people who have influenced directly or indirectly this dissertation: please, accept a collective thank for all you have taught me.

Finally, I should give a big thank to A Giridhar Rao, who patiently read the first three chapters of the final manuscript and improved its (and my) English – the revised version is on line since 10 January 2009 in my web space, permalink: federicogobbo.name.

Last but not least, I must record my gratitude, once and again, to my wife Francesca and our children, Susanna and Leonardo, for their patience while I was writing this dissertation.

F.G.
Ginostra/Milan, May 2008–January 2009

Part IV

Appendix

Summary / Riassunto / Resumo

English

In the field of computational linguistics, the representation of natural languages (NL) in formal grammar terms is a well-known issue and a wicked problem indeed. In fact, some formal grammars lack of lexical and semantic information representation, reducing the linguistic structure to syntax. On the other side, approaches like dependency grammars lose expressive power, under a computational point of view. Adpositional grammars (adgrams) are a novel grammar formalism that overcomes these limitations. Adgrams are based on Pennacchietti's research, who put together Brøndal's logic description of prepositions, Tesnerian notion of valence, Langacker's cognitive dichotomy trajector/landmark and Silvio Ceccato's pioneer work in the field on machine translation (MT). The result is a quasi-formal description of adpositions, being the junctors of language structure. This description is called adpositional space. Each adpositional space is made of four adpositional types: Plus ($\oplus$), Minus ($\ominus$), Slash ($\oslash$), Times ($\otimes$). The resulting structure is cognitively sound and formally inspiring, as adpositional trees (adtrees) can be built as special Porphyrian trees, going beyond the Tesnerian somehow fuzzy concept of dependency.

This dissertation puts Pennacchietti's work a step forward. In fact, here adgrams consider the ultimate unit of NLs being the morpheme, not the word, and therefore they offer a coherent theory of both morphology and syntax. Hence, the collocation phenomena are considered like zero morphemes. Moreover, a sharp distinction in the dictionary between the adpositional space, essentially made of closed morphemes, and the lexicon, made of open morphemes. Moving again from Tesnerian structural syntax, open morphemes always have a fundamental grammar character: stative (O), adjunctive (A, as stative modifiers), and verbal (I), circumstantial (E, as verbal modifiers). The Tesnierian approach is validated through Whorf's research results comparing grammars of typologically distant NLs.

The second part shows that adgrams are computable, as they can be implemented with a strong, robust formalism. A concrete instance of the formal model of adgrams is given through the quasi-natural language Esperanto (QNL, Lyons), showing the linguistic viability of the model. The formal model should be used appropriately in an ad hoc epistemological scenario, called 'the translation game', designed as a *Gedankenexperiment* à la Turing. A toy example is also given. In the third part the implementation is explained with all its details: the implementation of Esperanto is made in exactly 179 logic formulas, 56 of which are predicates. The formal model is promising to be generalised for any NLs, and how to do this was explained in various points of the third part. Finally, this dissertation shows that adgrams are a powerful NL grammar formalism which is at the same time cross-linguistic, cognitively grounded and formally robust and computationally sound.

Esperanto

En komputika lingvistiko, la formala reprezentado de la gramatiko de etnaj lingvoj estas ankoraŭ malferma kaj malfacila temo. Fakte, estas formalaj gramatikoj kiuj pretendas redukti la lingvan strukturon al sintakso, sen lekiska aŭ semantika reprezentado. Troviĝas ankaŭ aliroj leksike pli riĉaj, kiel ekzemple dependenco-gramatikoj, sed tiuj perdas esprimivon laŭ vidpunkto de la formaleco. Adpoziciaj gramatikoj (adgramoj) estas nova gramatika formalismo kiu preteras tiujn limbarojn. Ĝi baziĝas je intuicio de Pennacchietti, kiu kunmetis logikan priskribon de prepozicioj fare de Brøndal, tesneran nocion de valenco, dikotomaĵon altrafigilo/gvidosigno de Langacker, kaj pioniran laboron pri peraŭtomata tradukado fare de Silvio Ceccato. La rezulto estas preskaŭ-formalan priskribon de adpozicioj, kiuj estas la kunmetiloj de la lingva strukturo. Tiu ĉi priskribo nomiĝas adpozicia spaco (adspaco). La adpozicia spaco de ĉiu lingvo faratas de kvar adpoziciaj tipoj: Plus (⊕), Minus (⊖), Onigo (⊘), Obligo (⊗). La rezulta strukturo estas konive solida kaj interesa laŭ formala vidpunkto, ĉar adpozicia arbo (ad-arbo) povas konstruati kiel speciala porfira arbo, preterirante la nebulan koncepton de dependenco laŭ Tesnière.

Ĉi tiu doktora disertaĵo sukcesas fari kroman paŝon post la laboro de Pennacchietti. Fakte, ĉi tie adgramoj konsideras kiel lasta unuo de lingvoj la morfemon, ne la vorton, kaj sekve ili ofertas koheran teorion por morfologio kaj sintakso. Krome, faratas klara disiĝo en la vortaro inter adspaco, kies elementoj estas esence fermaj morfemoj, kaj leksiko, farata de malfermaj morfemoj. Denove ekde struktura sintakso fare de Tesnière, malfermaj morfemoj havas fundamentan gramatikan karakteron: nomiga (O), aldonebla (A, kiel modifanto de nomigiloj), kaj krome verba (I) kaj cirkonstanca (E, kiel modifanto de verboj). La tesnera aliro estas kontrolita per la rezultoj en komparo inter gramatikoj de tipologie foraj lingvoj fare de Whorf.

La dua parto montras, ke adgramoj estas komputeblaj, ĉar realigitaj per formalismo forta kaj firma. Krome, ekzemplo de la formala modelo estas proponita per Esperanto, tiel, ke la taŭgeco de la modelo lingvavidpunkte estu klara. Propre, la formala modelo estus uzata en ekkonteoria scenaro, ĝuste ĉizita laŭ la okazo. Tiu ĉi scenaro nomiĝas la 'ludo de la traduko', kaj ĝi estas mensa eksperimento, samkiel Turing. Oni donas eĉ etan ekzemplon. La tria parto montras la realigon en ĉiaj aspektoj: fakte la realigo de Esperanto faratas de entute 179 da logikaj formuloj, kaj, ene de tiu aro, 56 estas predikatoj kaj la resto reguloj. La formala modelo esperigas pri ĝeneraliga valido al ĉia ajn lingvo. Kiel fari tion klarigitas dise en la tria parto. La disertaĵo montras, ke adgramoj estas plenpova formalismo de lingvaj gramatikoj, ĉar ili estas samtempe lingve trairpovaj, konive grundaj kaj komputike belsonaj.

Italiano

Nel campo della linguistica computazionale, la rappresentazione formale della grammatica delle lingue storico-naturali è un argomento a tutt'oggi aperto e dibattuto. Infatti, certe grammatiche formali pretendono di ridurre la struttura del linguaggio alla sintassi, senza rappresentazioni lessicali o semantiche. D'altra parte, approcci lessicalmente piú ricchi, come le grammatiche della dipendenza, perdono in capacità espressiva da un punto di vista formale. Le grammatiche adposizionali (adgram) sono un formalismo grammaticale nuovo che supera questi limiti. Si basano sulle ricerche di Pennacchietti, che ha messo insieme la descrizione logica delle preposizioni di Brøndal, la nozione tesneriana di valenza, la dicotomia traiettoria/contrassegno di Langacker, e il lavoro pionieristico di Silvio Ceccato nel campo della traduzione automatica. Il risultato è una descrizione pseudo-formale delle adposizioni, vale a dire i giuntori della struttura linguistica. Questa descrizione viene detta spazio adposizionale (adspazio). Lo spazio adposizionale di ciascuna lingua è composto da quattro tipi adposizionali: Piú (⊕), Meno (⊖), Diviso (⊘), Per (⊗). La struttura risultante è cognitivamente solida e interessante da un punto di vista formale, perché un albero adposizionale (ad-albero) può essere costruito come un particolare albero porfiriano, andando oltre il concetto di dipendenza di Tesnière, che risulta nebuloso.

Questa tesi di dottorato fa un passo ulteriore sul lavoro di Pennacchietti. Infatti, qui le adgram considerano l'unità ultima delle lingue storico-naturali il morfema, non la parola, e conseguentemente offrono una teoria coerente di morfologia e sintassi. Difatti, i fenomeni di collocazione sono trattati come morfemi zero. Inoltre, viene operata una distinzione netta nel dizionario tra lo spazio adposizionale, formato essenzialmente da morfemi legati, e il lessico, fatto di morfemi liberi. Partendo ancora una volta dalla sintassi strutturale tesneriana, i morfemi liberi hanno un carattere grammaticale fondamentale: stativo (O), aggiuntivo (A, come modificatore di stativi), e inoltre verbale (I), e circumstanziale (E, come modificatore di verbi). L'approccio tesneriano è validato attraverso i risultati di ricerca in grammatica comparata di lingue tipologicamente distanti effettuati da Whorf.

La seconda parte mostra che le adgram sono computabili, in quanto implementate mediante un formalismo forte, solido. Viene fornita un'istanza concreta del modello formale attraverso la lingua quasi storico-naturale (Lyons) esperanto, che mostra l'appropriatezza linguistica del modello. Il modello formale dovrebbe essere usato appropriatamente in uno scenario epistemologico fatto ad hoc, chiamato 'il gioco della traduzione', ideato come un esperimento mentale, à la Turing. Viene anche fornito un esempio giocattolo. La terza parte mostra l'implementazione in tutti i suoi dettagli: l'implementazione dell'Esperanto è composta esattamente da 179 formule

logiche, 56 delle quali sono predicati, il resto regole. Il modello formale dà speranza di poter essere generalizzato per qualsiasi lingua storico-naturale, e come farlo viene spiegato in diversi punti della terza parte. In conclusione, la tesi mostra che le adgram sono un formalismo delle grammatiche delle lingue storico-naturali potente, perché sono al contempo trasversali linguisticamente, fondate cognitivamente, formalmente robuste e buone a un punto di vista computazionale.

Copyright notice

Pubblicato in Italia.

Università degli Studi dell'Insubria

Dipartimento di Informatica e Comunicazione (DICOM)

via Mazzini 5

IT-21100

Varese

Italy

Table 9.1: Translation of the most important keywords

English	**Esperanto**	**Italian**
adjunctive	aldonebla	aggiuntivo
circumstantial	cirkonstancaĵo	circostanziale
clause	subfrazero	clausola
landmark	gvidosigno	contrassegno
phrase	frazero	frase
sentence	frazo	periodo
stative	stativa	stativo
trajector	altrafigilo	triettoria
transferee	transigito	traslato
transferer	transigilo	traslatore
verbal	verbigila	verbante

Bibliography

A. J. Ades and Mark J. Steedman. On the order of words. *Linguistics and Philosophy*, (4):517–558, 1982. 32

Vilmos Ágel and Guta Rau, editors. *Dependenz und Valenz: Ein Internationales Handbuch Der Zeitgenössischen Forschung = Dependency and Valency: an International Handbook of Contemporary Research*. Walter de Gruyter GmbH & Co, Berlin, 2003. 147, 368

Hassan Aït-Kaci. *Warren's abstract machine: a tutorial reconstruction*. MIT Press, Cambridge, MA, 1991. 332

Akademio de Esperanto. Akademia vortaro. Web site. URL `http://akademio-de-esperanto.org/akademia_vortaro/`. Retrieved 12 November 2008. 15, 145, 153

ALPAC. Language and Machines. Technical report, Division of Behavioral Sciences. National Academy of Sciences, National Research Council, Washington, D.C., 1966. URL `http://www.mt-archive.info/ALPAC-1966.pdf`. Publication 1416. 93

Douglas Arnold, Lorna Balkan, Siety Meijer, R. Humphreys, and Louisa Sadler. *Machine Translation: an introductory Guide*. Blackwell, London, 1994. Available online. 132

John Austin. *Philosophical Papers*. Oxford University Press, Oxford, 1961. 33

E. Bach and R. T. Harms, editors. *Universals in Linguistic Theory*. Holt, Rinehart & Winston, New York, 1968. 93, 361

Richard Bandler and John Grinder. *The Structure of Magic*. Science & Behavior Books, Palo Alto, CA, 1975. 94

Peter Benson. *Comprehensive English-Esperanto Dictionary*. Esperanto League for North America, El Cerrito, 1995. 227

Eckhard Bick. Tagging and Parsing and Artificial Language: An Annotated Web-Corpus of Esperanto. In Davies et al. [2007]. 149

Detlev Blanke. *Internationale Plansprachen. Ein Einfürung*. Akademie Verlag, Berlin, 1985. 152

Detlev Blanke. Eugen Wüster (1898-1977). Leben und Werk. ein österreichischer Pionier der Informationgesellschaft. His Life and Work. An Austrian Pioneer of the Information Society. In *Terminology Science and Planned Languages*, volume 1, pages 133–168, Wien, 1998. International Information Centre for Terminology. Proceedings of the International Conference on Professional Communication and Knowledge Transfer. 152

Detlev Blanke. *Interlinguistische Beiträge. Zum Wesen und zur Funktion internationaler Plansprachen*. Peter Lang, Frankfurt am Main u. a., 2006. Sabine Fiedler (Hg.). 152

Burtea Bodgan, Josef Tropper, and Helen Younansardaroud, editors. *Studia Semitica et Semitohamitica. Festschrift für Rainer Voigt anläßlich seines 60. Geburtstages am 17. Januar 2004*. Ugarit Verlag, 2005. 32, 367

Franks Boers, Jeroen Darquennes, Koen Kerremans, and Rita Temmerman, editors. Newcastle upon Tyne, 2008. Cambridge Scholar Publishing. Proceedings of MACL. Brussells, Belgium, 2006, 8 – 10 February. 32, 362

George Boole. The calculus of logic. *The Cambridge and Dublin Mathematical Journal*, 3, 1848. URL `http://www.maths.tcd.ie/pub/HistMath/People/Boole/CalcLogic/`. Available online. 32

Pier Giorgio Borbone, Alessandro Mengozzi, and Mauro Tosco, editors. *Loquentes linguis. Studi linguistici e orientali in onore di/ Linguistic and Oriental Studies in Honour of/ Lingvistikaj kaj orientaj studoj honore al Fabrizio A. Pennacchietti*. Harassowitz Verlag, Wiesbaden, 2006. 142, 362, 369

R. Borsley and K. Borjars, editors. *Non-Transformational Syntax*. Blackwell, London, 2007. 32, 368

Carl B. Boyer. *A History of Mathematics*. John Wiley & Sons, Hoboken, NJ, 1968. 253

Umberto Broccatelli. *Vocabolario esperanto-italiano*. Cooperativa Editoriale Esperanto, Milano, 1991. 227

Viggo Brøndal. *Præpositionernes Theori*. Ejnar Munksgaard, Copenhagen, 1940. Italian translation by Silva editrice, Milan, 1967. 36, 92

Viggo Brøndal. *Teoria delle prepositioni*. Silva editrice, Milan, 1967. Italian translation of Brondal 1940. 92

Montagu Butler. Pri landonomoj. *The British Esperantist*, 1947. Republished in Löwenstein 2007. 199

Terrel Ward Bynum, editor. *Conceptual notation and related articles*. Oxford University Press, Oxford, 1972. English translation and edition of Frege 1879. 32

Luigi Cagni, editor. *La lingua di Ebla*, Napoli, 1981. Istituto Universitario Orientale di Napoli. 32, 366

Luigi Cagni, editor. *Il bilinguismo a Ebla*, Napoli, 1984. Istituto Universitario Orientale di Napoli. 32, 367

Chris Callison-Burch, Miles Osborne, and Philipp Koehn. Re-evaluating the Role of BLUE in Machine Translation Research. In Diana McCarthy and Shuly Wintner, editors, *Proceedings of EACL-2006. 11th Conference of the European Chapter of the Association for Computational Linguistics*, pages 249–256, Trento, Italy, 2006. University of Trento. 139

Jesús Cardeñosa, Alexander Gelbukh, and Edmundo Tovar. Universal networking language: Advances in theory and application. *Research on Computing Science*, 2005. Special issue. 132, 147

Ugo Cassina, editor. *Formulario Mathematico*. Cremonese, 1960. Reprint of Peano 1908. 32

Silvio Ceccato. La meccanizzazione delle attività umane superiori. *Civiltà delle macchine*, (4):1–4, 1961a. 92, 93

Silvio Ceccato. La traduzione nell'uomo e nella macchina. *Civiltà delle macchine*, (5):55–61, 1961b. 46, 92, 93

Silvio Ceccato. *Operational Linguistics and Translation*, pages 11–80. In , Ceccato [1961d], 1961c. 93

Silvio Ceccato, editor. *Linguistic Analysis and Programming for Mechanical Translation*. Feltrinelli editore / Gordon & Breach Publisher, Milano / New York, 1961d. 93, 359

André Cherpillod. Po. la diabla vorto. Eldonejo La Blanchetière, FR-72320 Cougenard, 1996. 220

André Cherpillod. La vortfarado en dek lecionoj. Speciale elpensita por la malgramatikemuloj. Eldonejo La Blanchetière, FR-72320 Cougenard, 2003. 199, 229, 253

André Cherpillod. Dore domilado solresol. historio, gramatiko, leksikono de la universala muzika lingvo solresol. Eldonejo La Blanchetière, FR-72320 Cougenard, 2008. 152

Noam Chomsky. *Syntactic Structures*. Mouton, The Hague, 1957. 23, 32, 33

Noam Chomsky. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA, 1965. 32

Thierry Coquand, Bengt Nordstrom, Jan M. Smith, and Bjorn Von Sydow. Type theory and programming. *EATCS*, 52:203–228, 1994. 332

Thomas T. Cormen, Charles E. Leiserson, and Ronald L. Rivest. *Introduction to algorithms*. MIT Press, Cambridge, MA, 1990. 332

Renato Corsetti, Maria A. Pinto, and Maria Tolomeo. Regularizing the regular: The phenomenon of overregularization in esperanto-speaking children. *Language Problems Language Planning*, 28:261–282, 2004. 153

Louis Couturat. *La logique de Leibniz d'apres des Documents Inedits*. Alcan, Paris, 1901. 32, 152

Louis Couturat. *Opuscules et fragments inédits de Leibniz*. Alcan, Paris, 1903. 32, 152

Louis Couturat and Leopold Leau. *Histoire de la langue universelle*. Hachette, Paris, 1903. 152

Louis Couturat and Leopold Leau. *Les nouvelles langues internationales*. Hachette, Paris, 1907. 152

William Croft and D. Alan Cruse. *Cognitive Linguistics*. Cambridge University Press, Cambridge, 2004. 33, 34, 92

Matthew Davies, Paul Rayson, Susan Hunston, and Pernilla Danielsson, editors. *Proc. of the Corpus Linguistics Conference CL2007*, 2007. University of Birmingham, UK. 149, 357

Martin Davis, editor. *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems And Computable Functions*. Raven Press, New York, 1965. 32

Edwin de Kock. *La Esperantaj landnomoj tamen kontentigaj*, pages 66–71. In , Löwenstein [2007], 2007. 229

Ferdinand De Saussure. *Introduzione al secondo corso di linguistica generale (1908 – 1909)*. Astrolabio, Roma, 1970. Editor: Robert Godel. Italian edition: Raffaele Simone. 32, 97

René De Saussure. *Fundamentaj reguloj de la vort-teorio*. Inko, 2003. Editor: Franko Luin. Original Edition: 1915. 229

Ralph Debusmann. *A declarative grammar formalism for dependency grammar*. PhD thesis, Universität des Saarlandes, 2001. 92

Gabriele Di Stefano and Vianello Renzo, editors. *Psicologia dello sviluppo e problemi educativi: Studi in onore di Guido Petter*. Giunti, Firenze, 2002. 228, 367

Joh. Dietterle, editor. *Originala verkaro*. Oriental-libro, Osaka, 1983. Reprint. Original edition: Ferdinand Hirt & Sohn Esperanto-fako, Leipzig 1929. 152, 227

Germain Dondelinger, Francoise Madray-Lesigne, and Jeannine Richard-Zappella, editors. *Actes du Colloque International*, 1995. Peeters Publishers. 32, 364

Michel Duc-Goninaz. *Nova Plena Ilustrita Vortaro de Esperanto*. Sennacieca Asocio Tutmonda, Paris, 2 edition, 2002. 146, 227

Helen S. Eaton. The educational value of an artificial language. *Modern Language Journal*, (12):87–94, 1927. 228

Umberto Eco. *La ricerca della lingua perfetta nella cultura europea*. Laterza, Bari, 1993. English translation published by Basil Blackwell, 1995. 32, 33, 93, 132, 152, 253

Sabine Fiedler. *Plansprache und Phraseologie*. Peter Lang, Frankfurt am Main, 1999. 144

Sabine Fiedler. Standardization and self-regulation in an international speech community: the case of esperanto. *Int'l. J. Soc. Lang.*, 177:67–90, 2006. 15, 145, 146

Sabine Fiedler and Haitao Liu, editors. *Studoj pri Interlingvistiko. Festlibro omaĝe al la 60-jariĝo de Detlev Blanke. Studien zur Interlinguistik. Festschrift für Detlev Blanke zum 60. Geburtstag*. Kava-Pech, Dobřichovice (Prago), 2001. 32, 229, 364

John C. Fillmore. *The case for case*, pages 1–88. In , Bach and Harms [1968], 1968. 93

John C. Fillmore. Frame semantics. *Linguistics in the Morning Calm*, pages 111–38, 1982. 33

John C. Fillmore. Frames and the semantics of understanding. *Quaderni di Semantica*, 0(6):222–53, 1985. 33

Peter G. Forster. *The Esperanto Movement*. Mouton, The Hague, 1982. 152

Helmar Frank and Reinhard Fössmeier. *AIS. La Akademio Internacia de la Sciencoj San Marino. Die Internationale Akademie der Wissenschaften San Marino*. Institut für Kybernetik, 2000. 152

Franco Furlan and Gaetano A. Lanzarone. *Prolog. Linguaggio e metodologia di programmazione logica.* Franco Angeli, Milano, 1988. 332

Jeremi Gishron. *Lingvo kaj religio – Studo pri la frua esperantismo kun speciala atento al L. L. Zamenhof*. Sivron, Jerusalem, 1986. 152

Ernst von Glaserfeld. *Silvio Ceccato and the Correlational Grammar*, pages 313–324. In , Hutchins [2001], 2001. Available on line. 93, 152

Christopher Gledhill. *The Grammar of Esperanto. A corpus-based description*. Lincom Europa, München, 1 edition, 1998. 145, 198, 208, 227, 229

Christopher Gledhill. *The Grammar of Esperanto. A corpus-based description*. Lincom Europa, München, 2 edition, 2001. 151, 227

Federico Gobbo. Il dilemma dell'esperanto. Master's thesis, Università degli Studi di Torino, 1998. URL `http://federicogobbo.name`. Available on line. 34, 103, 142

Federico Gobbo. The European Union's Need for an International Auxiliary Language. *Journal of Universal Language*, 6:1–28, March 2005a. 32, 142, 150, 152

Federico Gobbo. The digital way to spread conlangs. In Santiago Posteguillo, editor, *Language at work: language learning, discourse and translation studies in Internet*, volume 1, pages 45–53, Castellon, Spain, 2005b. Publicacions de la Universitat Jaume I. ICIL2005, Spain, 27 – 28 October. 152

Federico Gobbo. Planning languages and language planning: The contribution of interlinguistics to cross-cultural communication. In Boers et al. [2008], pages 24–60. Proceedings of MACL. Brussells, Belgium, 2006, 8 – 10 February. 13, 32, 142, 143, 152

Federico Gobbo. *L'esperanto e la traduzione automatica: Storia, risultati e prospettive esperantologiche dell'approccio statistico*, pages 319–327. In , Borbone et al. [2006], 2006. 142

Federico Gobbo. Pianificare il lessico scientifico internazionale: Peano e Wüster a confronto. Giuseppe Peano and his School between Mathematics Logic and Interlingua, October 2008b. International congress of study. 229

H. Herman Goldstine and John von Neumann. On the principles of large scale computing machines. In A. H. Taub, editor, *John von Neumann Collected Works*, volume V, pages 1–32. The Macmillan Co., New York, 1963. 332

Ma Gotoo, Tida Yamasaki, Sibayama Huĝimoto, and Usuim SAkaguchi, editors. *Aziaj kontribuoj al esperantologio. Aktoj de la 30-a Esperantologia Konferenco en la 92-a Universala Kongreso de Esperanto*, Rotterdam, The Netherlands, 2007. Universala Esperanto-Asocio. 228, 369

Algirdas J. Greimas. *Sémantique structurale*. Larousse, Paris, 1966. 33

E. Grosjean-Maupin, A. Esselin, S. Grenkamp-Kronfeld, and Gaston Waringhien. Plena vortaro de esperanto kun suplemento, 1956. 146, 227

Francesca Gualandri. The relationship between natural languages and formalisms, December 2008. Personal communication. 332

Reinhard Haupenthal, editor. *Naturalismo kaj skemismo. Korespondo inter Fernando de Diego kaj Karolo Piĉ kaj Reinhard Haupenthal*. Artur E. Iltis, Saarbrücken, 1987. 227

Adolf Holzhaus. *Doktoro kaj lingvo Esperanto*. Fondumo Esperanto, Helsinki, 1969. 152

Adolf Holzhaus. *Provo de gramatiko de novjuda lingvo kaj alvoko al la juda intelektularo*. Fondumo Esperanto, Helsinki, 1982. 152, 227

John N. Hutchins, editor. *Early years in machine translation*. John Benjamins, Amsterdam, 2001. 93, 362

John N. Hutchins and Evgenii Lovtskii. Petr Petrovich Troyanskii (1894-1950): A forgotten pioneer of mechanical translation. *Machine Translation*, 15(3):187–221, 2000. 146, 341

John N. Hutchins and Harold L. Somers. *An Introduction to Machine Translation*. Academic Press, London, 1992. Available on line. 132, 141, 146, 147, 152

Roger Imbert, Tibor Sekelij, Ivica Špoljarec, Spomenka Štimec, and Zlatko Tišljar. *Esperanto. Introduzione alla lingua internazionale. Enkonduko en la internacia lingvo*. Cooperativa Editoriale Esperanto, Milano, 1985. 227, 228

Kanzi Itô. *senlegenda biografio de L. L. Zamenhof*. Ludovikito, Kioto, 1982. 152

Kanzi Itô, editor. *Lingvaj respondoj de Ludoviko. Konsiloj kaj opinioj pri Esperanto*. Ludovikito, Kioto, 1990. 227, 228, 229

Kanzi Itô, editor. *unuaj libroj*. Ludovikito, Kioto, 1991. 146, 152

Kanzi Itô. *Historieto de Esperanto*. Ludovikito, Kioto, 1998. 152

Ray Jackendoff. *Semantic Structures*. MIT Press, Cambridge, MA, 1990. 33

Wim Jansen. *Woordvolgorde in het Esperanto. Normen, tallgebruik en universalia*. Landelijke Onderzoekschool Tallwetenschap, Utrecht, 2007. 156, 165, 216, 227, 253

Wim Jansen. *Naturaj vortordoj en Esperanto*. Universala Esperanto-Asocio, Rotterdam, The Netherlands, 2008. 227

Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. Prentics-Hall, Upper Saddle River, NJ, 2000. 32

Fred Karlsson, Atro Voutilainen, Heikkilä, and Arto Anttila. *Constraint Grammar. A Language-Indipendent System for Parsing Unrestricted Text*. Mouton de Gruyter, Berlin & New York, 1995. 148

J. Katz and J.A. Fodor. The structure of a semantic theory. *Language*, 39: 170–210, 1963. 97

Brian W. Kernighan and Dennis M. Ritchie. *The C Programming Language*. Prentice Hall Professional Technical Reference, 1988. ISBN 0131103709. 256

András Kornai. On the proper definition of information. In Terrell Ward Bynum, Mariacarla Calzarossa, Ivo de Lotto, and Simon Rogerson, editors, *Living, Working and Learning Beyond Technology*, pages 488–495, Mantua, 2008. University of Pavia. 98

Erich-Dieter Krause. *Großes Wörterbuch Esperanto-Deutsch*. Helmut Buske Verlag, Hamburg, 1999. 227

Erich-Dieter Krause. *Großes Wörterbuch Deutsch-Esperanto*. Helmut Buske Verlag, Hamburg, 2007. 227

Andreas Kück. Oficialaj lingvoelementoj de esperanto (oleo). Selbstverlag, Rechtenfleth, Germany, 2008. 153

Andy Künzli. *René de Saussure (1868-1943). Tragika sed grava esperantologo kaj interlingvisto el Svislando*. In , Fiedler and Liu [2001], 2001. available on line. 32, 229

George Lakoff. *Women, fire and dangerous things: what categories reveal about the mind*. Chicago University Press, Chicago, 1987. 33

Ronald W. Langacker. *Foundations of Cognitive Grammar*, volume 1. Stanford University Press, 1987. Theoretical Prerequisites. 33, 38

Ronald W. Langacker. Structural syntax: the view from cognitive grammar. In Dondelinger et al. [1995], pages 13–37. 32, 34

Andrew Large. *The Artificial Language Movement*. Basil Blackwell, Oxford, 1985. 152

Jouko Lindstedt. *A Man of Measure: Festschrift in Honour of Fred Karlsson*, volume 19, chapter Native Esperanto as a Test Case for Natural Language, pages 47–55. Turku Linguistic Association of Finland, 2006. Special Supplement to SKY Journal of Linguistics. 153

Ulrich Lins. *Die gefährliche Sprache. Die Verfolgung der Esperantisten unter Hitler und Stalin*. Bleicher, Gerlingen, 1988. 152

Anna Löwenstein, editor. *Rusoj loĝas en Rusujo. Landonomoj en Esperanto*. Federazione Esperantista Italiana, Milano, 2007. 229, 360

Erika Luciano and Claudia Silvia Roero. *Giuseppe Peano. Matematico e Maestro*. Dipartimento di Matematica, Università degli Studi di Torino, 2008. 152

John Lyons. *Semantics. Volumes I and II*. Cambridge University Press, Cambridge, 1977. 33

John Lyons. *Linguistic Semantics: an Introduction*. Cambridge University Press, Cambridge, 1995. 33

John Lyons. *Natural language and universal grammar. Essays in linguistic theory*, volume 1. Cambridge University Press, Cambridge, 2006. 139, 140, 150

N. Z. Maimon. *La kaŝita vivo de Zamenhof. Originalaj studoj*. Japana Esperanto-Instituto, Tokio, 1978. 152

Dan Maxwell and Klaus Schubert. *Metataxis in Practice: Dependency syntax for for multilingual machine translation*. DLT Distribution Language Traslation Series. Foris, Dortrecht, 1989. 147

James D. McCawley, editor. *Syntax and Semantics: Notes form the Linguistic Underground*. Academic Press, New York, 1976. 133, 367

Bruno Migliorini. *Manuale di Esperanto*. Cooperativa Editoriale Esperanto, Milano, 1995. 227, 228

Ken Miner. Noto pri esperantaj radikoj. *Lingva Kritiko*, 2008. 132

Carlo Minnaja. *Vocabolario Italiano-Esperanto*. Cooperativa Editoriale Esperanto, Milano, 1996. 227

Carlo Minnaja and L. G. Paccagnella. A part-of-speech tagger for esperanto oriented to mt, 2000. International Conference MT 2000. Machine Translation and multilingual Applications in the new Millennium. 149

Richard Montague. *The Proper Treatment of Quantification in Ordinary English*, pages 17–34. In , Portner and Partee [2002], 2002. URL `http://semantics.uchicago.edu/kennedy/classes/s08/semantics2/montague73.pdf`. Available on line. 230

Makoto Nagao. *A Framework of a Mechanical translation between Japanese and English by Analogy Principle*, chapter 11. Elsevier Science Publishers, 1984. Elithorn, A. and Banerji, R. Artificial and Human Intelligence. 141, 147, 341

Joakim Nivre. Dependency grammar and dependency parsing. Technical report, Växjö University, 2005. 35

Rihej Nomura. *Zamenhofa Ekzemplaro. La uzado de vortoj kaj esprimoj el la verko de D-ro L. L. Zamenhof*. Nagoja Esperanto-Centro, Nagoya, 1989. 227

Vincent Partington. *Implementation of an imperative programming language with backtracking*. PhD thesis, Faculty of Mathematics, Computer Science, Physics and Astronomy University of Amsterdam, Plantage Muidergracht 24, 1018 VT Amsterdam, 1997. 332

Lawrence C. Paulson. *Logic and computation: interactive proof with Cambridge LCF*. Cambridge University Press, Cambridge, 1987. 332

Lawrence C. Paulson. Isabelle: The next 700 theorem provers. In Pierluigi Odifreddi, editor, *Logic and Computer Science*, pages 361–386. Academic Press, 1990. 25, 332

Fabrizio A. Pennacchietti. Appunti per una storia comparata dei sistemi preposizionali semitici. *Annali dell'Istituto Orientale di Napoli*, pages 161–208, 1974. 32

Fabrizio A. Pennacchietti. La prepozicia sistemo de esperanto. *Esperantologiaj Kajeroj I*, pages 137–153, 1976. 32

Fabrizio A. Pennacchietti. Uno sguardo comparativo sul comparativo semitico. *Orientis Antiqvi Collectio*, 13:175–197, 1978. 32

Fabrizio A. Pennacchietti. Indicazioni preliminari sul sistema preposizionale dell'eblaita. In Cagni [1981], pages 291–319. 32, 94

Fabrizio A. Pennacchietti. I determinanti tipo quale e qualunque tra struttura superficiale e struttura sottostante. *Quaderni del seminario di iranistica, uralo-altaistica e causacasologia*, (19):347–368, May 1981b. La Bisaccia dello Sheikh. 32

Fabrizio A. Pennacchietti. Modi e forme del sintagma genitivale in semitico a partire dai testi di Ebla fino ai giorni nostri. In Cagni [1984], pages 291–319. 32

Fabrizio A. Pennacchietti. *Sull'etimologia e sul significato della preposizione araba àn*, pages 283–306. In , Bodgan et al. [2005], 2005. 32

Fabrizio A. Pennacchietti. Propono klasifiki la prepoziciojn de esperanto. In Wandel [2006], pages 68–83. Florence, Italy, 29 July – 5 August. 32, 148, 169, 170

Fabrizio A. Pennacchietti. Come classificare le preposizioni? una nuova proposta. *Quaderni del Laboratorio di Linguistica*, 6:1–18, 2006b. 32, 40, 92

Fabrizio A. Pennacchietti. *L'impiego di frasi pseudorelative come verbi finiti*, pages 133–148. In , Venier [2007], 2007. 94

Fabrizio A. Pennacchietti. The prepositional system of classic syriac and that of sureth. The 4^{th} Syriac Language Conference, November 2008. J/Simtha. Forthcoming. 32

W. Keith Percival. *On the Historical Source of Immediate-Constituent Analysis*, pages 229–242. In , McCawley [1976], 1976. Available online. 133

Robert Phillipson. *English-Only Europe? Language Policy Challenges*. Routledge, London, 2003. 137

Maria A. Pinto. *Bilinguismo e sviluppo metalinguistico*, pages 280–295. In , Di Stefano and Renzo [2002], 2002. 228

Claude Piron. *Esperanto: European or Asiatic Language?*, volume 22A. Universala Esperanto-Asocio, Rotterdam, The Netherlands, 1981. 228

Paul Portner and Barbara H. Partee, editors. *Formal Semantics: The Essential Readings*. Blackwell, London, 2002. 230, 366

Germano Proverbio and Anna Trocini Cerrina. *Elementi di sintassi strutturale*. Rosenberg & Sellier, Torino, 2001. 94

James Pustejovsky. *The Generative Lexicon*. MIT Press, Cambridge, MA, 1995. 33

Willard van Orman Quine. *Quiddities: An Intermittently Philosophical Dictionary*. Harvard University Press, Cambridge, MA, 1987. 132

Aarne Ranta. Grammatical framework: A type-theoretical grammar formalism. *The Journal of Functional Programming*, 14(2):145–189, 2004. 332

Aarne Ranta. Modular grammar engineering in gf. *Research on Language and Computation*, 5:133–158, 2007. Draft available online as pdf. Retrieved 25 November 2008. 332

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, NJ, 2002. 32

Marcos F. Sabarís, Jose L. R. Alonso, Dafonte C., and Arcay B. Multilingual authoring through an artificial language, 2001. URL `http://www.xente.mundo-r.com/utl/`. Available on line. 147, 148

Victor Sadler. *Working with Analogical Semantics: Disambiguation Techniques in DLT*. Number 5 in DLT. Foris Publications, Dordrecht, the Netherlands, 1989. 147

John E. Savage. *Models of Computation: Exploring the Power of Computing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1997. ISBN 0201895390. 332

Klaus Schubert. *Syntactic Tree structures in DLT*. DLT Distribution Language Traslation Series. Foris, Dortrecht, 1986. 147

Klaus Schubert. *Metataxis: Constrastive dependency syntax for multilingual machine translation*. Number 6 in DLT Distribution Language Traslation Series. Foris, Dortrecht, 1987. URL `http://www.mt-archive.info`. Full text available. Retrieved August 2008, 27. 92, 147, 335

Klaus Schubert. *Maschinelle Übersetzung mit Dependenzgrammatiken*, pages 1129–1157. In , Ágel and Rau [2003], 2003. 147

John R. Searle. *Expression and Meaning*. Cambridge University Press, Cambridge, 1979. 34

John R. Searle. Minds, brains and programs. *The Behavioral and Brain Sciences*, 3:417–24, 1980. 149, 150

John R. Searle. *The Rediscovery of the Mind*. Cambridge University Press, Cambridge, 1992. 34

Mark Steedman and Jason Baldridge. *Combinatory Categorial Grammar*. In , Borsley and Borjars [2007], 2007. Draft 5.0. To appear. 32, 332

Gabriele Stein. *A Dictionary of English Affixes – Their function and meaning*. Lincom Europa, München, 2007. 101

Leon Sterling and Ehud Shapiro. *The art of Prolog: advanced programming techniques*. MIT Press, Cambridge, MA, 2 edition, 1994. 332

Geoffrey H. Sutton. *Concise Encyclopedia of the Original Literature of Esperanto*. Mondial, New York, 2008. 144, 228, 250

István Szerdahelyi. *Eszperantó. Bevezetés és nyelvgyakorlatok Kézirat*. Tankönyvkiadól, Budapest, 3 edition, 1989. 16, 183, 228

John R. Taylor. *Linguistic Categorization*. Oxford University Press, Oxford, 1989. 32, 33, 92

John R. Taylor. *Cognitive Grammar*. Oxford University Press, Oxford, 2002. 27, 32, 33, 34, 93, 94, 101, 113, 132, 202, 204

Lucien Tesnière. *Éleménts de syntaxe structurale*. Editions Klincksieck, 8 rue de la Sorbonne, Paris, 1959. 32, 48, 54, 94, 103, 104, 132, 133, 147, 160, 227

Syintarô Tida. Kial multas lokaj adpozicioj en esperanto? In Gotoo et al. [2007]. 175, 228

Sebastiano Timpanaro. *La genesi del metodo del Lachmann*. Liviana Editrice, Padova, 1990. 93

Humphrey Tonkin, Grant Goodall, Ian Richmond, Mark Fettes, and Tim Reagan. Esperanto corpus available online to researchers. *Esperantic Studies*, (16):2, 2005. 149

Antonio Toral, Sergio Ferrández, and Andrés Montoyo. Eagles compliant tagset for the morphosyntactic tagging of esperanto. In Ruslan Mitkov, editor, *Recent Advances in Natural Language Processing*, Borovets, Bulgaria, 2005. Bulgarian Academy of Sciences. 149

Mauro Tosco. *Towards a geomoetry of adpositional systems: A preliminary investigation of Gawwada*, pages 695–702. In , Borbone et al. [2006], 2006. 46

Alan Turing. Computer Machinery and Intelligence. *Mind*, 59:433–460, 1950. 149

Hervé Vallon, editor. *Bibliographie seiner Veröffentlichungen. Esperantologie. Interlinguistik*. Harassowitz Verlag, Wiesbaden, 2006. Bilingual German-Esperanto. 152

Federica Venier, editor. *Relative e pseudorelative tra grammatica e testo*. Edizioni dell'Orso, 2007. 94, 367

John von Neumann. First draft of a report on the edvac. In N. Stern, editor, *From ENIAC to Univac: An Appraisal of the Eckert-Mauchly Computers*. Digital Press, Bedford, Massachusetts, 1981. 332

Amri Wandel, editor. *Internacia Kongresa Universitato (IKU) 2006*, Rotterdam, The Netherlands, 2006. Universala Esperanto-Asocio. Florence, Italy, 29 July – 5 August. 32, 367

Martin Warin. A comparison between a rule-based and a stochastic pos-tagger for esperanto. Technical report, C-uppsats i Allmän språkvetenskaop med inriktning på datorlingvistik, 2004. 149

Joseph Weizenbaum. Eliza – a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966. 32

Bertilo Wennergren. *Plena Manlibro de Esperanta Gramatiko*. Esperanto-Ligo por Norda Ameriko, El Cerrito, CA, 2005. 157, 159, 167, 168, 169, 171, 173, 211, 217, 227, 228, 253, 332

Benjamin Lee Whorf. Grammatical categories. *Language*, 21(1):1–11, 1945. URL `http://www.jstor.org/stable/410199`. JSTOR. Accessed: 16 June 2008; 09:22 CET. 100, 132

F. I. Wiener. Ŝajnprefiksoj. *Esperantologio*, 1(3):145–151, March 1951. 227

Anna Wierzbicka. *The Semantics of Grammar*. John Benjamins, Amsterdam, 1988. 93, 97

Dirk Willkommen. *Esperanto-Grammatik*. Helmut Buske Verlag, Hamburg, 2007. 165, 168, 227, 228

Niklaus Wirth. *Algorithms & data structures*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1986. 259, 332

Toon Witkam. Nova vojo al Automata Tradukado. Available on line, 2005. Talk presented at the Esperanto Universal Congress at Vilnius. 147

Seikô Yamasaki. *Enkonduko en la japanan*. Fonto, Chapecó, Brasil, 2000. 94, 228

Index